Technical Report 2603-VK8R

Validity Analyses for easyCBM® in Grades K-8: Reading

Gerald Tindal, PhD – University of Oregon

Sara McCaslin, PhD – mccaslinwordsmithing.wordpress.com

Published by

Behavioral Research and Teaching
University of Oregon • 175 Education
5262 University of Oregon • Eugene, OR 97403-5262
Phone: 541-346-3535 • Fax: 541-346-5689
http://brt.uoregon.edu

Note: This technical report was supported in part by Riverside Insight, the exclusive distributor of easyCBM®. The report does not reflect any endorsement by any of these organizations.

APA Reference: Tindal, G. (2026). *Validity Analyses for easyCBM® in Grades K-8: Reading (Technical Report # 2603-VK8R)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

[1] [1] For individual technical reports, see https://brtprojects.org: BRT Tech Reports / Search here…

Introduction

In the *Standards for Educational and Psychological Testing* (2014), "validity refers to the degree to which evidence and theory support the interpretation of test scores for proposed uses of the test. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests…It is the interpretations of the test for proposed uses that are evaluated, not the test scores themselves (p. 11)" (American Educational Research Association et al., 2014). The Standards emphasize a unified view of validity: diverse evidence is integrated into a coherent argument that links the construct to the score meaning, the uses, and the decisions made. The *Standards* describe several major sources of validity evidence. Evidence based on test content evaluates how well the tasks, items, and scoring align with the intended construct and domain. Evidence based on response processes examines the cognitive, behavioral, or rater processes engaged by examinees and scorers and whether they match the construct definition. Evidence based on internal structure evaluates the relationships among items, components, and scores (e.g., dimensionality, factor structure, invariance) and their consistency with the construct. Evidence based on relations to other variables examines expected convergent, discriminant, criterion, and group-difference relations. Evidence based on consequences of testing considers intended and unintended outcomes, including fairness, to the extent that consequences provide information about score meaning and use. Together, these sources justify interpretations and identify limitations.

In this technical report, four evidential sources are used to demonstrate validity for using easyCBM® early reading (K-2) measures and later measures in reading for students in Grades 3-8: (a) criterion related to demonstrate the reading constructs are related to other acceptable measures of reading. (b) regression analyses to support predictions of performance and accounted variance in other acceptable measures of reading, (c) internal structures to reflect uniform constructs in the measures, and (d) classification accuracy to use the measures for indicating risk of potential learning problems. are reported for the. These sources of evidence are applied to the following measures.

Table 1. Grade-Level Benchmark Measures by Assessment Period

Grade Level	Fall Benchmark	Winter Benchmark	Spring Benchmark
K	Letter Names Letter Sounds Phoneme Segmenting	Letter Sounds Phoneme Segmenting Word Reading Fluency	Letter Sounds Phoneme Segmenting Word Reading Fluency
1	Letter Sounds Phoneme Segmenting Word Reading Fluency	Letter Sounds Word Reading Fluency Passage Reading Fluency	Letter Sounds Word Reading Fluency Passage Reading Fluency
2	Passage Reading Fluency Vocabulary Proficient Reading	Passage Reading Fluency Vocabulary Proficient Reading	Passage Reading Fluency Vocabulary Proficient Reading

In Grades 3-8, the measures are Vocabulary, Passage Reading Fluency, and Comprehension (Basic and Proficient Reading).

Validity, as defined in the Standards for Educational and Psychological Testing (2014), refers to the degree to which evidence and theory support the interpretations of test scores for their intended uses. The unified framework emphasizes multiple sources of evidence, including evidence based on **relations to other variables** (criterion-related), internal structure, response processes, and consequences of testing. Across the easyCBM® Reading Technical Reports, a coherent and cumulative validity argument emerges grounded primarily in relations to external criteria, construct coherence, and classification accuracy within MTSS frameworks.

Evidence based on **relations to other variables** is consistently strong. Across grades K–2, early literacy measures (Letter Names, Letter Sounds, Phoneme Segmenting, Word Reading Fluency) demonstrate moderate to strong correlations with external benchmarks such as DIBELS, TOWRE, and SAT-10. Regression models frequently account for substantial variance in standardized outcomes, with combined predictor models explaining between 35% and over 70% of variance depending on grade and season. **Predictive validity** strengthens when criterion measures closely align with the easyCBM® construct (e.g., WRF predicting SAT-10 Word Reading). In Grades 3–8, Passage Reading Fluency (PRF), Multiple-Choice Reading Comprehension (MCRC), and Vocabulary show moderate

predictive relationships with state assessments (e.g., OAKS, MSP, SBAS), with
correlations often ranging from .57 to .68 for PRF and strong concurrent relations (.88–.95) with DIBELS ORF.

Evidence based on **internal structure** further supports construct validity. Confirmatory factor analyses generally support multi-factor structures distinguishing early literacy, fluency, and comprehension in lower grades, and coherent reading constructs in upper grades. Factor loadings for fluency and foundational skills are consistently strong, while comprehension indicators demonstrate moderate associations, occasionally influenced by ceiling effects. Rasch modeling across reports confirms item functioning, alternate-form equivalence, and appropriate coverage of the ability continuum.

Classification accuracy analyses provide consequential validity evidence related to screening decisions. ROC analyses demonstrate acceptable to strong AUC values (often .75–.95), with sensitivity prioritized to minimize false negatives in RTI contexts. Cross-validation studies in Oregon and Washington show stable cut scores across split samples, reinforcing defensibility of benchmark decisions within state-specific contexts.

Longitudinal and growth evidence supports the interpretation of reading fluency as a robust general indicator of reading competence, particularly in elementary grades. Growth modeling demonstrates sensitivity to instructional effects, and updated ORF norms confirm stability and modest national gains over time.

The following conclusions can be made from the validity studies with easyCBM® measures.

1. Criterion validity is present with other external measures, when using correlational analyses among single measures. Furthermore, the correlations among the various measures themselves are quite high.
2. When using regression analyses, the various measures work well together in predicting performance on other, external measures, far better than when used as single measures.
3. Although the National Center on Intensive Intervention (NCII) requires these relations to be documented for individual measures, this use is contrary to effective practices, where multiple measures are typically recommended.
4. The classification analyses indicate that most measures have sufficient area under the curve (AUC) and sensitivity. However, specificity is lower, indicating overidentification of students needing service. Three caveats need to be considered: (a) the analyses were conducted with the 40th Percentile Rank (PR) as the cut-off, which may be too high; (b) the cost of making false-positive and false-negative decisions need to be considered; (c) identification of risk on a singular measure reflects poor practice.
5. In another section of this technical report, multiple screening measures are used to identify students at risk of problems in learning to read; this mosaic of performance can then also be to identify effective progress monitoring measures.

Collectively, the integrated findings across technical reports align with the Standards' unified view of validity. Evidence from criterion relations, internal structure, classification accuracy, and growth modeling converges to support the interpretation of easyCBM® Reading scores for screening, progress monitoring, and instructional decision-making within MTSS systems. While no single statistic is sufficient, the cumulative evidence forms a coherent and technically defensible validity argument across Grades K–8.

Note 1: Two important caveats need to be considered in this and other summary reports on easyCM®: (a) Many reports reflect studies with multiple indicators of technical adequacy with both reliability and criterion validity investigated. In a similar manner, both criterion validity and classification accuracy may be reported in the same TR. (b) The reports typically include students from multiple grades with overlap across the TRs. We have attempted, however, to address both caveats by reporting on the main type of technical adequacy so that the suite of these reports are grouped as documents focusing on test development, alignment, reliability, and validity. Furthermore, the general sequence is to report results from earlier to later grades.

Note 2: All tables and figures in this summary are examples of those presented in full within the individual Technical Reports (TR) but are not exhaustive, just illustrative.

Summary of Technical Report 1003: Technical Adequacy of the easyCBM® Primary-Level Reading Measures (Grades K–1), 2009–2010 Version (Lai et al., 2010).

Methods

This technical adequacy report documents validity evidence for easyCBM® primary-level reading measures used in Kindergarten and Grade 1. Measures include foundational early literacy tasks (e.g., Letter Names, Letter Sounds, Phoneme Segmenting), and early reading fluency indicators (e.g., Word Reading Fluency; and grade-appropriate fluency measures). Validity evidence includes relations between easyCBM® measures and external achievement outcomes, notably SAT-10/SESAT Word Reading. Regression models are used to evaluate predictive validity (both single-predictor and multi-predictor), reporting R and R^2 to quantify the strength of association with the standardized criterion.

Results (quotes from pages 11-14 of the Technical Report)

Concurrent validity evidence. The spring model for Grade K was significant, $F(3,179) = 158.80$, $p < .05$, and accounted for 73% of the variance in SAT-10. The spring Word Reading measure had the highest coefficient, $b = 1.98$, $t(189) = 11.49$, $p < .05$, and uniquely explained 20.16% of the variance in the regression model. The descriptive statistics and regression model summaries for Grade K are presented in Tables 5-8. The spring model for Grade 1 was significant, $F(1,178) = 234.35$, $p < .05$, accounting for 57.8% of the variance in SAT-10. The spring Passage Reading Fluency benchmark had a coefficient, $b = .71$, $t(185) = 15.31$, $p < .05$, and uniquely explained 56.78% of the variance. The descriptive statistics and regression model summaries for Grade 1 are presented in Tables 9-12.

Predictive validity evidence. The fall model for Grade K was significant, $F(3,158) = 69.72$, $p < .05$, and accounted for 57% of the variance in SAT-10. The fall LS benchmark had the highest coefficient, $b = 2.48$, $t(2013) = 6.09$, $p < .05$, and uniquely explained 10.11% of the variance. The winter model for Grade K was significant, $F(3,169) = 112.11$, $p < .05$, and accounted for 66.6% of the variance in SAT-10. The winter WRF benchmark had the highest coefficient, $b = 2.48$, $t(1972) = 8.27$, $p < .05$, and uniquely explained 13.54% of the variance.

The fall and winter model for Grade K was significant, $F(6,155) = 86.18$, $p < .05$, and accounted for 76.9% of the variance in SAT-10. The winter WRF benchmark had the highest coefficient, $b = 4.02$, $t(1972) = 8.09$, $p < .05$, and uniquely explained 9.73% of the variance. Descriptive statistics and regression model summaries for Grade K are presented in Tables 13-22. The fall model for Grade 1 was significant, $F(3,155) = 48.18$, $p < .05$, accounting for 48.3% of the variance in SAT-10. The fall WRF benchmark had the highest coefficient, $b = .82$, $t(2179) = 2.54$, $p < .05$, and uniquely explained 2.16% of the variance. The winter model for Grade 1 was significant, $F(3,173) = 73.64$, $p < .05$, and accounted for 56.1% of the variance in SAT-10. The winter LS benchmark had the highest coefficient, $b = .92$, $t(2195) = 5.16$, $p < .05$, and uniquely explained 6.76% of the variance. The model in which we included both fall and winter results for Grade 1 was significant, $F(6,152) = 33.87$, $p < .05$, and accounted for 57.2% of the variance in SAT-10. The winter LS benchmark had the highest coefficient, $b = 1.03$, $t(2195) = 4.47$, $p < .05$, and uniquely explained 5.62% of the variance. Descriptive statistics and regression model summaries for Grade 1 are presented in Tables 23-32.

Table 2. Example Regression Predictors in Reading

Table 6

Kindergarten Model Summary

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.853[a]	.727	.722	26.704	.727	158.798	3	179	.000

a. Predictors: (Constant), Spr10WRF, Spr10Seg, Spr10LS

Explanation of Model Summaries for Regression Analyses (see Table 3 above)

The regression "Model Summary" shows how well the predictor in the footnote predicts the dependent variable used in the analysis (the outcome name is not displayed in the snippet). The row labeled "1" is Model 1, which includes three predictor (easyCBM® = Spr10WRF, Spr10Seg and Spr10LS) and a constant (intercept).

R = .853. R is the multiple correlation between observed outcome scores and the scores predicted by the regression equation. With only one predictor, it is essentially the absolute value of the Pearson correlation between Spr10PRF and the outcome. An R of .853 indicates a strong positive linear relationship: students with higher scores on the measures (Spr10WRF, Spr10Seg and Spr10LS) tend to have higher predicted outcome scores.

R Square = .727. R^2 (the coefficient of determination) is the proportion of variance in the outcome explained by the model. Here, the three measures (Spr10WRF, Spr10Seg and Spr10LS) accounts for 72.7% of the variability in the Kindergarten outcome—substantial explanatory power for a single predictor.

Adjusted R Square = .722. Adjusted R^2 corrects R^2 for sample size and number of predictors to reduce optimism from fitting a model to a particular dataset. The adjustment is a function of the number of predictors and the sample size. The adjustment is small (.727 → .722). This suggests the explained variance is not driven by overfitting and would likely be similar in comparable samples.

Std. Error of the Estimate = 26.704. This is the standard deviation of the residuals (prediction errors) in the outcome's units. It represents the typical distance between a student's observed outcome score and the model's predicted score. Interpreting its magnitude depends on the outcome scale, but as a rule of thumb, smaller values indicate more precise prediction.

R Square Change = .727. The "Change Statistics" section is produced for hierarchical regression. R^2 Change is the increment in explained variance when the predictor block is entered. Because the measures are the first and only block, the change equals the total R^2.

F Change = 158.798; df1 = 3; df2 = 179. This F test evaluates whether adding the three measures improve model fit beyond an intercept-only model (i.e., whether the R^2 increase is greater than 0). df1 equals the number of predictors entered at this step (1). df2 is the residual degrees of freedom (N − predictors − 1). With df2 = 179, the implied sample size is about N = 180. The very large F value indicates the improvement in fit is enormous relative to residual error.

Sig. F Change = .000. This is the p-value for the F test (SPSS prints .000 when $p < .001$). It indicates the R^2 increase from adding the three measures is statistically significant. In practical terms, the three measures are highly reliable predictor of the Kindergarten outcome in this sample, and the combination of high R and high R^2 indicates both statistical and practical importance.

Overall, the three measures (Spr10WRF, Spr10Seg and Spr10LS) capture many differences on the outcome, but prediction is not perfect: unexplained variance remains ($1 - R^2 = .273$), reflecting other skills, instructional factors, and error.

Table 3. Example Variance Explained Output with easyCBM® Early Reading Measures

Table 10

Grade 1 Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.754[a]	.568	.566	27.202	.568	234.347	1	178	.000

a. Predictors: (Constant), Spr10PRF

Table 14

Kindergarten Model Summary (Fall easyCBM and SAT-10)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.755[a]	.570	.561	33.557	.570	69.715	3	158	.000

a. Predictors: (Constant), Fall09WRF, Fall09Seg, Fall09LS

Table 17

Kindergarten Model Summary (Winter easyCBM and SAT-10)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.816[a]	.666	.660	29.564	.666	112.114	3	169	.000

a. Predictors: (Constant), Wint10WRF, Wint10Seg, Wint10LS

Table 20

Kindergarten Model Summary (Fall + Winter easyCBM and SAT-10)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.877[a]	.769	.760	24.802	.769	86.178	6	155	.000

a. Predictors: (Constant), Wint10WRF, Wint10Seg, Fall09Seg, Wint10LS, Fall09LS, Fall09WRF

Table 24

Grade 1 Model Summary (Fall easyCBM and SAT-10)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.695[a]	.483	.473	29.985	.483	48.184	3	155	.000

a. Predictors: (Constant), Fall09PRF, Fall09LS, Fall09WRF

Table 27

Grade 1 Model Summary (Winter easyCBM and SAT-10)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.749[a]	.561	.553	27.597	.561	73.641	3	173	.000

a. Predictors: (Constant), Wint10PRF, Wint10LS, Wint10WRF

Table 30

Grade 1 Model Summary (Fall + Winter easyCBM and SAT-10)

Model					Change Statistics				
	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.756[a]	.572	.555	27.536	.572	33.865	6	152	.000

a. Predictors: (Constant), Wint10PRF, Wint10LS, Fall09LS, Fall09PRF, Wint10WRF, Fall09WRF

Construct validity evidence. For grade K, the model fit for the hypothesized 1-factor model for all seasons was good, with CFI indices ranging from .997 to .999, TLI indices ranging from .992 to .997 and RMSEA indices ranging from .028 to .047. Descriptive statistics and full CFA model fit indices are presented in Tables 33-34. For grade 1, CFI and TLI indices suggested good model fit for the hypothesized 1-factor model for all seasons with CFI indices ranging from .978 to .988 and TLI indices ranging from .979 to .982. RMSEA, on the other hand, while indicating a fair model fit for the fall model suggested poor model fit for winter and spring models, with RMSEA indices ranging from .106 to .136. Descriptive statistics and full CFA model fit indices are presented in Tables 35-36.

Figure 1. Example Construct Validity Structure

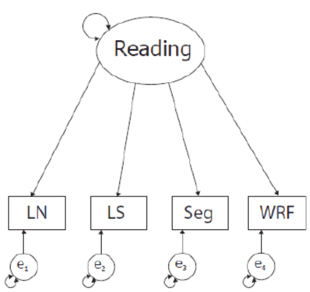

Figure 1. Hypothesized 1-Factor Model for easyCBM Reading Kindergarten

Note. LN = Letter Naming; LS = Letter Sounds; Seg = Phoneme Segmenting; WRF = Word Reading Fluency

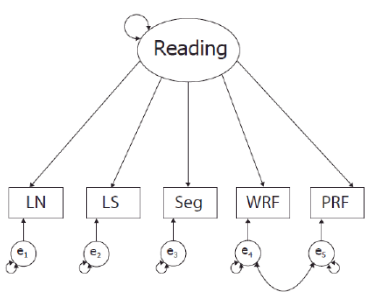

Figure 2. Hypothesized 1-Factor Model for easyCBM Reading Grade 1

Note. LN = Letter Naming; LS = Letter Sounds; Seg = Phoneme Segmenting; WRF = Word Reading Fluency; PRF = Passage Reading Fluency

Reference

Lai, C.-F., Nese, J. F. T., Jamgochian, E. M., Alonzo, J., & Tindal, G. (2010). *Technical A\adequacy of the easyCBM® primary-level reading measures (Grades K–1), 2009–2010 Version (Technical Report #1003).* Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1004: Technical Adequacy of the easyCBM® Grade 2 Reading Measures (Jamgochian et al., 2010).

Methods

This report summarizes technical adequacy for Grade 2 easyCBM® reading measures, including criterion-related validity evidence. Measures include grade-appropriate reading indicators used for screening and progress monitoring. Criterion-related evidence is examined by relating easyCBM® scores to SAT-10 Word Reading outcomes using regression model summaries (R and R^2), evaluating how well earlier-season easyCBM® measures predict spring standardized reading performance.

Results

Concurrent validity. None of the grade 2 easyCBM® measures were significantly correlated with the SAT-10. Correlations ranged from .036 - .052 (Table 8). Regression analyses were conducted for each of the measures separately and combined (easyCBM® model). Individually, the measures were weak predictors of SAT-10 scores. In the combined model, MCRC was the best predictor.

Predictive validity. Fall WRF and PRF were significantly correlated with SAT-10 (.154 [$p < .05$] and .194 [$p < .01$], respectively). Winter WRF and PRF were also significantly correlated with SAT-10 (.151 [$p < .05$] and .221 [$p < .01$], respectively). Neither Fall nor Winter MCRC was significantly correlated with SAT-10. Regression analyses were conducted for each of the measures separately and combined by season. For Fall, the individual measures were weak predictors of SAT-10. In the combined model, WRF and MCRC were better predictors than PRF. For Winter, individually, MCRC was the best predictor of SAT-10. In the combined model, MCRC was also the best predictor, WRF was moderate, and PRF was negative. Descriptive statistics are presented in Table 7 and regression model summaries are presented in Tables 17 –24 (Fall) and 25 - 32 (Winter).

Table 4. Correlations Among easyCBMs® in Reading and Criterion Measures

Table 8

Grade 2 Correlation Matrix (easyCBM® N = 3675; SAT-10 N = 205)

		Fall 2009			Winter 2010			Spring 2010			
		WRF	PRF	MCRC	WRF	PRF	MCRC	WRF	PRF	MCRC	SAT-10
Fall 2009											
WRF	Pearson Correlation		.999**	.936**	.913**	.912**	.907**	.801**	.836**	.880**	.154*
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.028
PRF	Pearson Correlation			.934**	.912**	.912**	.906**	.799**	.835**	.877**	.194**
	Sig. (2-tailed)			.000	.000	.000	.000	.000	.000	.000	.005
MCRC	Pearson Correlation				.916**	.914**	.917**	.803**	.838**	.866**	.070
	Sig. (2-tailed)				.000	.000	.000	.000	.000	.000	.318
Winter 2010											
WRF	Pearson Correlation					.998**	.971**	.848**	.891**	.905**	.151*
	Sig. (2-tailed)					.000	.000	.000	.000	.000	.031
PRF	Pearson Correlation						.971**	.848**	.892**	.904**	.221**
	Sig. (2-tailed)						.000	.000	.000	.000	.001
MCRC	Pearson Correlation							.834**	.872**	.901**	-.081
	Sig. (2-tailed)							.000	.000	.000	.251
Spring 2010											
WRF	Pearson Correlation								.885**	.889**	.047
	Sig. (2-tailed)								.000	.000	.502
PRF	Pearson Correlation									.925**	.052
	Sig. (2-tailed)									.000	.458
MCRC	Pearson Correlation										.036
	Sig. (2-tailed)										.608

**Correlation is significant at the 0.01 level. *Correlation is significant at the 0.05 level.

Reference

Jamgochian, E., Park, B. J., Nese, J. F. T., Lai, C.-F., Sáez, L., Anderson, D., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM® grade 2 reading measures (Technical Report #1004)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1005: Technical Adequacy of the easyCBM® Reading Measures (Grades 3–7), 2009–2010 Version (Sáez et al., 2010).

Methods

This report examines the technical adequacy of the easyCBM® reading assessments across grades 3–8. Measures included Passage Reading Fluency (PRF), Multiple-Choice Reading Comprehension (MCRC), and Vocabulary. Data were drawn from large-scale district administrations during the 2009–2010 school year. Reliability analyses included internal consistency for MCRC and Vocabulary, alternate form reliability, and growth slope reliability estimated using hierarchical linear modeling. Validity analyses included concurrent and predictive correlations with state reading assessments, regression modeling by season, and confirmatory factor analyses evaluating the structural coherence of constructs across grades. Descriptive statistics and quartile-based analyses allowed examination of differential performance patterns. Growth reliability was analyzed conditional on baseline performance.

Results

Internal consistency coefficients for MCRC and Vocabulary were generally moderate to strong. Growth slope reliability varied across grades and quartiles, with higher reliability in lower-performing groups. PRF demonstrated consistent predictive validity with state outcomes, particularly in fall and winter administrations. **Criterion validity** using regression analyses indicated that seasonal models explained meaningful variance in state test performance, with spring models often showing the strongest concurrent validity. Confirmatory factor analyses supported a multidimensional structure separating fluency, comprehension, and vocabulary constructs.
Overall, the measures demonstrated acceptable reliability and **strong criterion-related validity across grades**, supporting their use for RTI-based screening and monitoring.

Table 5. Illustrative Table from Technical Report 1005

Table 386
Grade 6 Fall Easy CBM Scores Predicting Spring OAKS Reading Performance

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change
1	.764[a]	.584	.582	6.098	.584	434.920	3	931	.000

a. Predictors: (Constant), Fall09Voc, Fall09MCRC, Fall09PRF

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	199.301	.884		225.565	.000	197.567	201.035					
	Fall09PRF	.074	.006	.315	11.475	.000	.061	.086	.645	.352	.243	.594	1.684
	Fall09MCRC	.401	.069	.154	5.820	.000	.266	.536	.551	.187	.123	.638	1.568
	Fall09Voc	.886	.057	.427	15.473	.000	.773	.998	.693	.452	.327	.589	1.699

a. Dependent Variable: OAKSRdgTot

Reference

Sáez, L., Park, B. J., Nese, J. F. T., Jamgochian, E., Lai, C. F., Anderson, D., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM reading measures (Grades 3–8), 2009–2010 version (Technical Report No. 1005)*. University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1309: easyCBM® Reading Criterion-Related Validity Evidence: Grades K–1 (Lai et al., 2013b).

TR1309 reports a criterion-related validity study examining how easyCBM® early reading measures relate to other well-established literacy assessments for Kindergarten and Grade 1. Participants were recruited from a convenience sample of three elementary schools in one Oregon district that used easyCBM® as part of an RTI model. The study included 214 Kindergarten students and 227 Grade 1 students. Data were collected over three weeks in February 2013. Fifteen University of Oregon undergraduate/graduate data collectors were trained by a licensed school psychologist in standardized administration and scoring; fidelity checks included supervision, double-check scoring, and random verification.

Methods

easyCBM® measures included Phoneme Segmenting (PS), Letter Names (LN), Letter Sounds (LS), and Word Reading Fluency (WRF), each administered individually for 60 seconds using standardized protocols. Comparator measures were selected to target similar constructs and included DIBELS (Kindergarten: Initial Sound Fluency, Phoneme Segmenting Fluency, Letter Naming Fluency; Grade 1: Phoneme Segmenting Fluency, Letter Naming Fluency, Nonsense Word Fluency), the CTOPP Elision subtest (phonological awareness), and the TOWRE-2 Sight Word Efficiency subtest (word reading). To reduce order and form effects, easyCBM® and comparator measures were assembled into a booklet and administered in one of two counterbalanced rotations. Data preparation coded reasons for missingness, and correlation analyses were conducted. Because linearity/normality assumptions were often not met, the primary statistic was Spearman's rho.

Results

Overall, easyCBM® measures showed a wide range of relations with comparator measures, with the strongest evidence when content, format, and administration procedures were closely aligned. In Kindergarten, descriptive statistics indicated broad score ranges on each measure; only 188 students completed the TOWRE due to scheduling constraints. Correlations showed a high relation between easyCBM® Phoneme Segmenting and DIBELS Phoneme Segmenting Fluency (rs = .85), and a high relation between easyCBM® Letter Names and DIBELS Letter Naming Fluency (rs = .86). easyCBM® Letter Sounds had a moderate relation with DIBELS Initial Sound Fluency (rs = .55). easyCBM® Word Reading Fluency was moderately high with TOWRE Sight Word Efficiency (rs = .79). In contrast, easyCBM® Phoneme Segmenting showed a low relation with CTOPP Elision (rs = .39).

In Grade 1, all measures were administered to 214 students except TOWRE (n = 210 in the table). easyCBM® Phoneme Segmenting was moderately to highly related to DIBELS Phoneme Segmenting Fluency (rs = .75), and easyCBM® Letter Names was highly related to DIBELS Letter Naming Fluency (rs = .80). easyCBM® Letter Sounds showed a moderate relation with DIBELS Nonsense Word Fluency (rs = .58). easyCBM® Word Reading Fluency showed a very high relation with TOWRE Sight Word Efficiency (rs = .95). However, the association between easyCBM® Phoneme Segmenting and CTOPP Elision was essentially zero and non-significant in Grade 1 (rs = .05).

The discussion emphasizes that the strongest correlations likely reflect close correspondence in content and procedures (e.g., easyCBM® PS with DIBELS PSF; easyCBM® LN with DIBELS LNF). Moderate relations for Letter Sounds are interpreted as consistent with related but not identical tasks (e.g., picture-based initial sound identification or decoding unfamiliar letter strings). The weak relations with CTOPP Elision were unexpected and are attributed to using only a single CTOPP subtest rather than the recommended phonological awareness composite; the authors recommend including additional CTOPP subtests in future studies. Overall, the report concludes that easyCBM® PS and LN show strong criterion evidence with DIBELS, WRF shows strong criterion evidence with TOWRE, and LS shows moderate criterion evidence with DIBELS measures.

Summary. In the screen shot below, for Kindergarten and Grade 1:

1. Phoneme Segmenting and Letter Names from easyCBM® correlate highly with their corresponding measures from Dibels.
2. Word Reading from easyCBM® correlates highly with a similar measure from TOWRE.

Table 6. Correlations Among easyCBMs® Reading with DIBELS, CTOPP, and TOWRE

Table 9

Spearman's Rho Rank Correlation Results – Kindergarten

easyCBM® Measures		DIBELS Phoneme Segmenting Fluency	DIBELS Letter Naming Fluency	DIBELS Initial Sound Fluency	CTOPP Elision	TOWRE Sight Word Efficiency
		Comparator Measures				
Phoneme Segmenting	r_s	.85**	-	-	.39**	-
	n	227	-	-	222	-
Letter Names	r_s	-	.86**	-	-	-
	n	-	227	-	-	-
Letter Sounds	r_s	-	-	.55**	-	-
	n	-	-	225	-	-
Word Reading Fluency	r_s	-	-	-	-	.79**
	n	-	-	-	-	188

Note. ISF = Initial Sound Fluency, PSF = Phoneme Segmenting Fluency, LNF = Letter Naming Fluency, PS = phoneme segmenting, LN = Letter names, LS = Letter sounds, WRF = Word reading fluency; **. Correlation is significant at the 0.01 level (2-tailed).

Table 10

Spearman's Rho Rank Correlation Results – Grade 1

easyCBM® Measures		DIBELS Phoneme Segmenting Fluency	DIBELS Letter Naming Fluency	DIBELS Nonsense Word Fluency	CTOPP Elision	TOWRE Sight Word Efficiency
		Comparator Measures				
Phoneme Segmenting	r_s	.75**	-	-	0.05	-
	n	214	-	-	214	-
Letter Names	r_s	-	.80**	-	-	-
	n	-	214	-	-	-
Letter Sounds	r_s	-	-	.58**	-	-
	n	-	-	214	-	-
Word Reading Fluency	r_s	-	-	-	-	.95**
	n	-	-	-	-	210

**. Correlation is significant at the 0.01 level (2-tailed).

Note. PSF = Phoneme Segmenting Fluency, LNF = Letter Naming Fluency, NWF = Nonsense Word Fluency, PS = phoneme segmenting, LN = Letter names, LS = Letter sounds, WRF = Word reading fluency; **. Correlation is significant at the 0.01 level (2-tailed).

Reference

Lai, C.-F., Alonzo, J., & Tindal, G. (2013). *easyCBM® reading criterion-related validity evidence: Grades K–1 (Technical Report #1309).* Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 0922: Technical Adequacy of the easyCBM® Reading Measures (Grades 3–8), 2009–2010 Version (Tindal et al., 2009b)

This technical report investigates how student characteristics influence growth in passage reading fluency (PRF) across grades 3 through 8 using data from the easyCBM® assessment system. The primary goal of the study was to analyze patterns of reading fluency development over time and determine how demographic and educational factors relate to both initial reading performance and growth rates.

Methods

The study analyzed districtwide data collected during the 2009 school year. Students in grades 3 through 8 were assessed on passage reading fluency during three benchmark periods: fall, winter, and spring. Passage reading fluency was measured as the number of words read correctly per minute during timed oral reading passages administered through the easyCBM® system. Sample sizes were large at each grade level, typically including approximately one thousand students per grade. Student characteristics examined in the analysis included gender, race or ethnicity, economic disadvantage, special education status, Title I participation, and identification as historically low achieving.

To analyze reading growth across the school year, the researchers used hierarchical linear modeling (HLM). This statistical approach allowed the analysis to account for the nested structure of the data, where multiple reading assessments were collected for each student across time. In the model, time served as the first level of analysis, representing the repeated fall, winter, and spring assessments, while student characteristics were included at the second level. The modeling approach made it possible to examine both the intercept, representing initial reading fluency levels, and the slope, representing the rate of reading growth over time. By including student characteristics in the model, the analysis estimated how these variables influenced both starting performance and growth patterns.

Results

Results indicated that student characteristics significantly influenced both initial reading fluency scores and the rate of fluency growth across the school year. Several demographic variables were associated with lower starting levels of reading fluency, including being male, belonging to a racial or ethnic minority group, being economically disadvantaged, and receiving either special education or Title I services. These characteristics also influenced growth rates in some grade levels, though the strength and direction of the relationships varied across grades.

The hierarchical models showed that the relationship between initial reading performance and subsequent growth differed by grade level. In some grades, students who began with higher fluency scores showed slower growth rates, while in other grades the pattern differed. Overall, the results demonstrate that reading fluency growth is influenced by multiple student characteristics and that these relationships vary across grade levels.

Table 7. Illustrative Table from Technical Report 0922

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	185.379007	1.658585	111.769	1212	0.000
GENDER_N, G01	9.875223	1.961743	5.034	1212	0.000
ETHNICIT, G02	-9.950520	2.321850	-4.286	1212	0.000
ECONDIS, G03	-12.662480	2.193945	-5.772	1212	0.000
SPECED, G04	-36.176189	3.304185	-10.949	1212	0.000
For ORDER slope, B1					
INTRCPT2, G10	-0.726933	0.115489	-6.294	1212	0.000
GENDER_N, G11	-0.275166	0.144826	-1.900	1212	0.057
ETHNICIT, G12	0.280367	0.187889	1.492	1212	0.136
ECONDIS, G13	0.064711	0.160073	0.404	1212	0.686
SPECED, G14	-0.300663	0.248768	-1.209	1212	0.227

Reference

Tindal, G., Nese, J. F., & Alonzo, J. (2009b). *Hierarchical linear modeling of passage reading fluency growth as a function of student characteristics (Technical Report # 0922)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1801: Classification Accuracy of the easyCBM® Kindergarten – Grade 2 Reading Measures (Alonzo & Anderson, 2018a).

Methods

TR1801 reports classification-accuracy analyses designed to identify screening cut scores for easyCBM® literacy measures in Kindergarten through Grade 2. Data came from one small Pacific Northwest school district and included all enrolled students who were present during benchmark windows from fall 2016 through spring 2017 (Kindergarten n=288; Grade 1 n=315; Grade 2 n=326). Assessments were administered by district personnel (teachers or instructional assistants) using district benchmark protocols. Prior to testing, staff completed easyCBM's online training modules and demonstrated proficiency for each measure they administered.

Measures varied by grade and season. Kindergarten included Phoneme Segmenting (SEG), Letter Names (LN; fall only), Letter Sounds (LS), and Word Reading Fluency (WRF; winter/spring). Grade 1 included SEG (fall), LS, WRF, and Passage Reading Fluency (PRF; winter/spring). Grade 2 included PRF plus group-administered Multiple-Choice Reading Comprehension (MCRC) and Vocabulary (VOC) in fall, winter, and spring. Classification accuracy was evaluated with ROC analyses reporting Area Under the Curve (AUC) and candidate thresholds. Threshold selection followed a decision rule: keep sensitivity ≥0.80 and maximize specificity; if specificity exceeded 0.90, then keep sensitivity ≥0.90 and maximize specificity; ties were resolved by selecting the cut maximizing sensitivity. Outcomes were defined using spring 'benchmark' criteria anchored to national norms: Kindergarten outcome = spring LS 50th percentile raw score (32); Grade 1 outcome = spring WRF 50th percentile raw score (49); Grade 2 outcome = spring PRF 50th percentile raw score (101).

Results

Table 8 summarizes ROC results (AUC, selected threshold, sensitivity, specificity) by grade, measure, and season. Except for Kindergarten fall SEG, all measures showed acceptable-to-strong diagnostic accuracy (AUCs generally in the mid-.70s to high-.90s), and the selected thresholds met the report's sensitivity-first decision rules.

Table 8. Example Classification Accuracy Output for Early Reading Measures

Measure	Grade	Season	AUC	Threshold	Sensitivity	Specificity
LN	K	fall	0.76	6.5	0.82	0.66
LS	K	fall	0.68	1.5	0.8	0.49
SEG*	K	fall	0.37	---	1	0
LS	K	win	0.87	18.5	0.82	0.81
SEG	K	win	0.75	40.5	0.81	0.45
SEG	K	spr	0.85	49.5	0.80	0.77
SEG	1	fall	0.73	46.5	0.81	0.36
LS	1	fall	0.84	33.5	0.81	0.75
WRF	1	fall	0.95	11.5	0.84	0.89
LS	1	win	0.81	45.5	0.81	0.63
WRF	1	win	0.98	22.5	0.9	0.93
MCRC	2	fall	0.76	7.5	0.83	0.57
VOC	2	fall	0.82	10.5	0.85	0.64
PRF	2	fall	0.92	54.5	0.81	0.85
MCRC	2	spr	0.81	10.5	0.84	0.61
VOC	2	win	0.82	11.5	0.82	0.59
MCRC	2	win	0.81	9.5	0.81	0.63
PRF	2	win	0.94	84.5	0.9	0.74
VOC	2	spr	0.85	11.5	0.85	0.64

Overall, the report concludes that easyCBM® K–2 literacy measures support strong screening decisions when cut scores are selected to prioritize sensitivity while maintaining adequate specificity, with the notable exception of fall Kindergarten SEG.

Reference

Alonzo, J., & Anderson, D. (2018). *Classification accuracy of the easyCBM® Kindergarten – Grade 2 Reading Measures (Summary of Technical Report # 1801).* Eugene. OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1805: Supplementary Report on easyCBM® Grade K–2 Literacy Measures: Classification Accuracy and Predictive Validity (Follow-Up to TR 1403) (Alonzo & Anderson, 2018a).

Methods

This supplementary report reanalyzed a year-long K–1 dataset to provide additional indices and confidence intervals for classification accuracy and predictive validity. The study used fall easyCBM® alternate forms of Letter Names, Letter Sounds, Phoneme Segmenting, and Word Reading Fluency. External criteria were spring Stanford Achievement Test (SESAT/SAT-10) subtests appropriate to grade (e.g., kindergarten Sounds and Letters; Kindergarten Word Reading; Grade 1 word-related outcomes). Analyses relied on ROC methodology and area under the curve (AUC) to evaluate how well fall easyCBM® scores classified students relative to criterion benchmark cut points.

Results

For Kindergarten, AUC values varied across measures and alternate forms, but many fell in ranges commonly interpreted as acceptable to good for screening. **Letter Names** forms showed AUCs spanning roughly .69 to .85 (average ~.77) for predicting SAT-10 Sounds and Letters performance. **Letter Sounds** forms showed a narrower band (about .71 to .78; average ~.76). **Phoneme Segmenting** generally produced lower AUCs (roughly .62 to .71; average ~.66), consistent with a more specific construct and/or greater distance from some criteria. **Word Reading Fluency** forms in kindergarten also showed acceptable AUCs (about .71 to .78; average ~.76). Interpreting these values: AUC near .70 is often treated as minimally acceptable discrimination for universal screening; around .80 is good; values closer to .60 indicate limited discrimination for that pairing. The report also highlights variability across forms and criterion pairings, supporting the practical recommendation to use multiple indicators (and professional judgment) when decisions have higher stakes than routine screening.

See Table 9 on the next page for all coefficients displayed by Grade (Kindergarten, Grade 1, and Grade 2).

easyCBM® Reading Measures and SAT-10 Form: Sounds and Letters

- Area Under the Curve (AUC)
- Sensitivity
- Specificity

Kindergarten Measures:

- Letter Names
- Letter Sounds
- Phoneme Segmenting
- Word Reading

Grade 1 Measures

- Phoneme Segmenting
- Word Reading

Grade 2 Measures

- Passage Reading

Note that the **SAT 10** criterion measure changes with the easyCBM® grade level measures and includes

- Sounds & Letters
- Word Reading
- Vocabulary

Table 9. AUC with 40th PR on Test Forms of SAT-10

1.Kindergarten Letter Names –> SAT-10 Sounds & Letters

SAT-10 Form: Sounds and Letters	8	9	10	11	12	13	14	15	16	17
Area Under the Curve (AUC)	0.77	0.69	0.85	0.77	0.69	0.85	0.77	0.69	0.85	0.77
Sensitivity	0.74	0.8	0.8	0.8	0.84	0.71	0.76	0.82	0.79	0.84
Specificity	0.63	0.56	0.56	0.59	0.61	0.57	0.5	0.55	0.57	0.61

2. Kindergarten Letter Sounds –> SAT-10 Sounds & Letters

SAT-10 Form: Sounds and Letters	8	9	10	11	12	13	14	15	16	17
Area Under the Curve (AUC)	0.76	0.78	0.75	0.77	0.76	0.71	0.75	0.77	0.74	0.76
Sensitivity	0.76	0.84	0.82	0.78	0.8	0.69	0.82	0.84	0.74	0.86
Specificity	0.66	0.5	0.52	0.57	0.53	0.62	0.51	0.52	0.57	0.57

3. Kindergarten Phoneme Segmenting –> SAT-10 Sounds & Letters

SAT-10 Form: Sounds and Letters	5	7	7	8	10	11	12	13	14	15
Area Under the Curve (AUC)	0.7	0.64	0.69	0.65	0.65	0.64	0.62	0.71	0.67	0.66
Sensitivity	0.67	0.76	0.83	0.74	0.77	0.65	0.77	0.88	0.81	0.74
Specificity	0.57	0.47	0.48	0.5	0.49	0.59	0.47	0.5	0.52	0.48

4. Kindergarten Phoneme Segmenting –> SAT-10 Word Reading

SAT-10 Form: Word Reading	5	7	8	10	11	12	13	14	15
Area Under the Curve (AUC)	0.71	0.6	0.66	0.66	0.6	0.72	0.64	0.68	0.66
Sensitivity	0.85	0.83	0.74	0.81	0.9	0.92	0.96	0.8	0.77
Specificity	0.43	0.31	0.33	0.31	0.44	0.32	0.33	0.36	0.3

5. Kindergarten Word Reading Fluency –> SAT-10 Sounds & Letters

SAT-10 Form: Sounds and Letters	8	9	10	11	12	13	14	15	16	17
Area Under the Curve (AUC)	0.76	0.78	0.75	0.77	0.76	0.71	0.75	0.77	0.74	0.76
Sensitivity	0.85	0.83	0.74	0.81	0.9	0.92	0.96	0.8	0.77	0.77
Specificity	0.43	0.31	0.33	0.31	0.44	0.32	0.33	0.36	0.3	0.58

6. Grade 1 Phoneme Segmenting –> SAT-10 Sounds & Letters and Word Reading

SAT-10 Form: Sounds Letters (SL) Word Reading (WR)	5 SL	5WR	12WR	12WR	15WR	15WR	16WR	16WR	17WR	17WR
Area Under the Curve (AUC)	0.57	0.59	0.6	0.47	0.57	0.54	0.63	0.53	0.7	0.63
Sensitivity	0.6	0.65	0.8	0.65	0.38	0.53	0.75	0.53	0.53	0.53
Specificity	0.33	0.56	0.67	0.51	0.46	0.49	0.5	0.49	0.49	0.49

7. Grade 1 Word Reading Fluency –> SAT-10 Sounds & Letters

SAT-10 Form: Sounds and Letters	10	16
Area Under the Curve (AUC)	0.6	0.77
Sensitivity	0.8	0.75
Specificity	0.67	0.5

8. Grade 2 Passage Reading Fluency –> SAT-10 Vocabulary

SAT-10 Form: Vocabulary	5	9	10	15	17
Area Under the Curve (AUC)	0.83	0.86	0.86	0.85	0.86
Sensitivity	0.92	0.96	0.96	0.94	0.97
Specificity	0.43	0.37	0.38	0.33	0.39

Interpreting these values: AUC near .70 is often treated as minimally acceptable discrimination for universal screening; around .80 is good; values closer to .60 indicate limited discrimination for that pairing. The report also highlights variability across forms and criterion pairings, supporting the practical recommendation to use multiple indicators (and professional judgment) when decisions have higher stakes than routine screening.

Reference

Alonzo, J., & Anderson, D. (2018). *Supplementary report on easyCBM® grade K–2 literacy measures: classification accuracy and predictive validity, a follow-up to Technical Report 1403 (Technical Report #1805).* Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1806: Supplementary Report on easyCBM® PRF Measures: A Follow-Up to Previous Technical Reports (Alonzo & Anderson, 2018c).

Methods

This supplementary technical report re-analyzed previously published easyCBM® Passage Reading Fluency (PRF) data in response to requests for additional statistical detail, particularly confidence intervals around classification, reliability, and validity estimates. The report is organized into three primary evidence categories: classification accuracy, reliability, and criterion-related validity.

Criterion validity was investigated in addition to **classification accuracy** analyses, using the Smarter Balanced English Language Arts (SBAS ELA) assessment served as the independent criterion measure. Students in Grades 3–8 from Pacific Northwest districts were included (school year 2014–2015). The 40th percentile of the easyCBM® National Norms was used as the benchmark cut score, reflecting updated accountability expectations under SBAS. Receiver Operating Characteristic (ROC) analyses were conducted in R to compute sensitivity, specificity, false positive and false negative rates, positive and negative predictive power, overall classification rate, Area Under the Curve (AUC), and associated 95% confidence intervals.

Reliability analyses included test-retest reliability (same form, one week apart), alternate-form reliability (different passages), and Generalizability Theory (G-Theory) studies. Samples were drawn from three public elementary schools (Grades 1–5). G-Theory analyses used a fully crossed two-facet design (persons × forms × occasions) to estimate variance components and compute G-coefficients and phi coefficients.

Criterion-related validity evidence was examined using two studies. Study 1 evaluated predictive validity via linear regression between easyCBM® PRF and SBAS ELA scores. Study 2 examined concurrent validity via bivariate correlations between easyCBM® PRF and DIBELS ORF.

Results

Predictive validity coefficients between PRF and SBAS ELA ranged from .57 to .68 across Grades 3–8, indicating moderate positive relationships. Concurrent validity correlations between PRF and DIBELS ORF were very strong (.88–.95), confirming that the two measures assess the same underlying oral reading fluency construct.
This supplementary analysis strengthens the technical argument for easyCBM® PRF by providing confidence intervals and expanded reliability evidence. The findings indicate:

- Strong alternate-form and test-retest reliability
- High generalizability across forms and occasions
- Moderate predictive validity with SBAS ELA
- Very strong concurrent validity with DIBELS ORF

Together, these results provide robust technical evidence supporting the use of easyCBM® PRF for screening and progress monitoring within accountability-aligned MTSS frameworks.

Table 10. Illustrative Table from Technical Report 1806

Table 7
Criterion-Related Validity Evidence

Type of Validity	Grade	Criterion	n	Coefficient	95% Confidence Interval*: Lower Bound	95% Confidence Interval*: Upper Bound
Predictive	3	SBAS English Language Arts	1303	0.67	0.63	0.71
Predictive	4	SBAS English Language Arts	1520	0.64	0.60	0.68
Predictive	5	SBAS English Language Arts	1539	0.68	0.64	0.71
Predictive	6	SBAS English Language Arts	1467	0.61	0.57	0.65
Predictive	7	SBAS English Language Arts	1415	0.62	0.58	0.66
Predictive	8	SBAS English Language Arts	1475	0.57	0.53	0.61
Predictive	3	SBAS English Language Arts	1280	0.67	0.63	0.71
Predictive	4	SBAS English Language Arts	1489	0.63	0.59	0.67
Predictive	5	SBAS English Language Arts	1575	0.68	0.64	0.71
Predictive	6	SBAS English Language Arts	1494	0.63	0.59	0.67

Reference

Alonzo, J., & Anderson, D. (2018). *Supplementary report on easyCBM® PRF measures: A follow-up to previous technical reports (Technical Report No. 1806).* Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1807: Supplementary Report on easyCBM® MCRC Measures: A Follow-up to Previous Technical Reports (Alonzo & Anderson, 2018b).

Methods

This technical report presents supplementary analyses of the easyCBM® Multiple Choice Reading Comprehension (MCRC) measures to provide additional evidence regarding classification accuracy, reliability, and criterion-related validity. The samples consisted of large-scale datasets of students in Grades 2–8 drawn primarily from school districts in the Pacific Northwest. Classification accuracy analyses used statewide Smarter Balanced English Language Arts (SBAS ELA) scores as the independent criterion measure. The analytic samples for classification accuracy included more than 26,000 to 34,000 students per grade for Grades 3–8 in the 2014–2015 school year. For reliability analyses, fall and winter benchmark data from the 2013–2014 school year were used, including over 20,000 students per grade in many cases. The easyCBM® MCRC assessments were administered in benchmark windows: fall, winter, and spring.

Classification analyses used a cut score corresponding to the 40th percentile from easyCBM® national norms to identify students at risk for reading difficulties. Statistical analyses were conducted in R and included calculation of sensitivity, specificity, positive and negative predictive power, overall classification accuracy, and area under the ROC curve (AUC) with confidence intervals. Reliability was examined using Cronbach's alpha and split-half reliability. Criterion-related validity was evaluated through bivariate correlations and linear regression analyses relating easyCBM® MCRC scores to SBAS ELA performance.

Findings

The classification accuracy analyses demonstrated that easyCBM® MCRC measures provide useful screening information for predicting student performance on the Smarter Balanced ELA assessment. In fall benchmark assessments, overall classification accuracy ranged from approximately .70 to .75 across Grades 3–8. Sensitivity values ranged from about .55 to .62, indicating that the assessment identified most students who would later perform below proficiency on the criterion assessment. Specificity values were consistently high, ranging from approximately .83 to .91 across grades. Area Under the Curve (AUC) estimates ranged from approximately .79 to .84, indicating strong classification performance for a universal screening tool. Winter benchmark results produced similar findings, with classification accuracy between .63 and .78 and AUC values between approximately .75 and .84.

Spring benchmark analyses showed slightly improved predictive performance for several grades. AUC estimates ranged from approximately .80 to .85, while specificity values were often above .80 and reached as high as .93 in some grades. Overall classification rates remained stable around .70 to .75. These results suggest that the MCRC measures provide increasingly accurate information about student reading comprehension performance as the academic year progresses.

Internal consistency reliability estimates were generally strong across grades. Cronbach's alpha values ranged from approximately .66 to .79 across Grades 2–8. Split-half reliability coefficients ranged from approximately .60 to .76, indicating moderate to strong internal consistency for the timed comprehension assessments. Reliability analyses conducted across student subgroups—including general education students, students receiving special education services, and demographic subgroups—showed similarly strong reliability coefficients.

Criterion-related validity analyses showed moderate to strong correlations between easyCBM® MCRC scores and SBAS ELA scores. Predictive validity correlations for fall MCRC scores predicting spring SBAS scores were approximately .62 to .63 in Grade 3. Concurrent validity correlations between spring MCRC and SBAS scores across Grades 3–8 generally ranged from approximately .54 to .68. These findings indicate that the easyCBM® MCRC assessments capture important aspects of reading comprehension and meaningfully predict performance on an independent statewide assessment of English Language Arts.

Table 11. Main Findings Summary for Technical Report 1807

Evidence Type	Key Results	Interpretation
Classification Accuracy	AUC ≈ .79–.85; classification rate ≈ .70–.75	MCRC effectively identifies students at risk for reading difficulty
Reliability	Cronbach's α ≈ .66–.79; split-half ≈ .60–.76	Moderate to strong internal consistency across grades
Criterion Validity	Correlations with SBAS ELA ≈ .54–.68	MCRC scores strongly relate to external measures of reading achievement

Reference

Alonzo, J., & Anderson, D. (2018). *Supplementary report on easyCBM® MCRC measures: A follow-up to previous technical reports (Technical Report No. 1807)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1403: easyCBM® Kindergarten Beginning Reading Measures: Alternate Form Reliability and Criterion Validity With the SAT-10 (Wray et al., 2014).

Methods

This report evaluated alternate form reliability (addressed in another technical report) and criterion validity for Kindergarten easyCBM® beginning reading measures administered as brief, individually administered fluency assessments. Measures included Letter Names Fluency, Letter Sounds Fluency, Phoneme Segmenting Fluency, and Word Reading Fluency. For criterion validity, easyCBM® scores were compared to spring Stanford Achievement Test (SAT-10/SESAT) subtests aligned to early literacy outcomes (e.g., Sounds and Letters; Word Reading). Analyses emphasized correlations and regression models predicting

SAT-10 outcomes from fall easyCBM® measures, providing evidence of how well early easyCBM® scores relate to later standardized achievement.

Results

Criterion-related validity was examined using regression models relating easyCBM® measures to SESAT/SAT-10 outcomes. In kindergarten, easyCBM® LN, LS, and PS jointly explained about 35%–40% of the variance in SAT-10 Sounds & Letters across time points, and about 48%–58% of the variance in SAT-10 Word Reading. LN was frequently a significant predictor, LS was often significant for SAT-10 Word Reading, and PS contributed inconsistently. In Grade 1, easyCBM® LN, LS, PS, and WRF explained about 14%–32% of SAT-10 Word Study Skills, with WRF consistently the strongest predictor (unique variance commonly in the teens to low 20s). When the criterion was restricted to word reading (easyCBM® WRF later in the year), the model explained about 49%–56% of the variance. The report concludes that the battery provides meaningful evidence of alternate-form reliability and criterion validity, with stronger prediction when the outcome is closely aligned to word reading.

Quotes from Tech Report (page 7 from the Technical Report)

Kindergarten. Results for the criterion validity of the kindergarten easyCBM® measures can be found in tables 18 and 19. The first model regressed easyCBM® LN, LS, and PS from time 1 to time 4 on SAT-10 Sounds and Letters. The overall variance explained by the model across time ranged from 35%---40%. For all time points, LN scores were a significant predictor of Sounds and Letters performance. LS scores were significant predictors only at Time 3 (February) and PS was a significant predictor for all time points except Time 3. The unique variance in performance explained by LN ranged from .16%---.64%. LS and PS both explained .03% of the unique variance across time. The second model regressed the same independent variables on the SAT-10 Word Reading measure. The overall model explained 48%---58% of the variance of the dependent variable. LN was a significant predictor for all time points except Time 2 (December). LS scores were significant for all time points, while PS was a non-significant predictor across all the time points. The variance of SAT-10 Word Reading uniquely explained by LN ranged from .01%---.81% and the variance uniquely explained by LS ranged from 16% – 64%.

Grade 1. Results for the criterion validity study of the grade 1 easyCBM® measures can be found in tables 20 and 21. The first model regressed easyCBM® LN, LS, PS, and WRF from time 1 to time 4 upon SAT---10 WSS. The variance explained by the whole model at each time point varied from 14%---32%. WRF was a significant predictor of SAT---10 WSS for all time points, and PS was significant for Time 4 (April) only. LN and LS were not significant predictors of SAT---10 WSS at any time point. The variance uniquely explained by PS at Time 4 was 2.0%, and the unique variance explained by WRF ranged from 13.0%---21.2%. The second model used the same independent variables, but substituted easyCBM® WRF at Time 4 for SAT---10 WSS. The variance explained by the model across time ranged from 49%---56%. LN performance was a significant predictor of easyCBM® WRF performance at Time 4 for all time points, and LS was significant for Times 2 (December) and 3 (February). PS was not a significant predictor of WRF performance at Time 4 at any time point. The variance in easyCBM® WRF score at Time 4 uniquely explained by LN ranged from 14.4%---41.0%, and the variance in easyCBM® WRF performance at Time 4 uniquely explained by LS performance ranged from 1.4 % – 2.3%.

Table 12. Example Regression Values with easyCBM® in Early Reading Measures

Table 18

Grade Kindergarten Regression Models Predicting SAT-10 Sounds and Letters

	Time 1			Time 2			Time 3			Time 4		
easyCBM © measures	*B*	*β*	Part	*B*	*β*	Part	*B*	*β*	Part	*B*	*β*	Part
Letter names	0.88*	0.34*	.04	0.98*	0.35*	.04	0.90*	0.33*	.04	1.13**	0.42**	.08
Letter sounds	0.61	0.16	.01	0.74	0.20	.01	0.86*	0.28*	.03	0.47	0.16	.01
Phoneme segmenting	0.54*	0.20*	.03	0.52*	0.19*	.03	0.33	0.11	.01	0.59*	0.17*	.03
R^2		.35			.39			.39			.40	
N		124			141			151			160	

Note. *p < .05, **p < .001, Part = Semipartial correlations.

Table Explanation. The tables report four concurrent regression models (Times 1–4) predicting SAT-10 Sounds and Letters from various easyCBM® measures. Standardized betas (β) show the relative strength of each predictor after all measures are in the model; larger |β| indicates a stronger unique association with the outcome, holding the other predictors constant. For example, in Table 18, across times, Letter Names is the most consistent predictor (β ≈ .33–.42) and is statistically significant at every time point, with its strongest effect at Time 4 (β = .42, p < .001). Letter Sounds is weaker early (β ≈ .16–.20, ns) but becomes a meaningful unique predictor at Time 3 (β = .28, p < .05). Phoneme Segmentation shows modest, significant effects at Times 1, 2, and 4 (β ≈ .17–.20), but not at Time 3.

The Part column gives the semi partial correlation for each predictor—i.e., the correlation between the outcome and the portion of that predictor not shared with the other predictors. Squaring Part yields the unique variance explained. For example in Table 18, Letter Names at Time 4 has Part = .08, so it uniquely explains about .006 (0.6%) of outcome variance beyond the other measures, despite a comparatively large β. Model R² values (.35–.40) indicate that the three easyCBM® measures together explain 35–40% of SAT-10 variance, with sample size increasing from N = 124 to 160.

Table 13. Example Regression Model Predictors between easyCBM® and SAT-10

Table 19

Grade Kindergarten Regression Models Predicting SAT-10 Word Reading

	Time 1			Time 2			Time 3			Time 4		
easyCBM © measures	*B*	*β*	Part	*B*	*β*	Part	*B*	*β*	Part	*B*	*β*	Part
Letter names	1.62**	0.49**	.09	0.64	0.19	.01	1.18**	0.36**	.05	0.96*	0.30*	.04
Letter sounds	1.46*	0.30*	.04	2.33**	0.52**	.08	1.44**	0.39**	.06	1.45**	0.42**	.08
Phoneme segmenting	0.02	0.01	.00	0.14	0.05	.00	0.19	0.05	.00	0.26	0.06	.00
R^2		.58			.49			.53			.48	
N		123			138			150			156	

Note. *p < .05, **p < .001, Part = Semipartial correlations.

Table 20

Grade 1 Regression Models Predicting SAT-10 Word Study Skills

easyCBM © measures	Time 1			Time 2			Time 3			Time 4		
	B	*β*	Part	*B*	*β*	Part	*B*	*β*	Part	*B*	*β*	Part
Letter names	0.09	0.03	.16	-0.09	-0.03	-.02	-0.37	-0.13	-.08	-0.33	-0.12	-.08
Letter sounds	0.34	0.09	.08	0.61	0.18	.12	0.15	0.05	.03	0.04	0.01	.01
Phoneme segmenting	0.10	0.03	.07	0.01	0.00	.00	0.64	0.14	.13	0.64*	0.15*	.14
Word reading fluency	1.04**	0.43**	.36	1.01**	0.45**	.38	1.15**	0.54**	.45	1.23**	0.61**	.46
R^2		.14			.28			.28			.32	
N		148			153			149			161	

Note. **p* < .05, ***p* < .001, Part = Semipartial Correlation.

Table 21

Grade 1 Regression Models Predicting the easyCBM© Word Reading Fluency measure at Time 5

easyCBM © measures	Time 1			Time 2			Time 3			Time 4		
	B	*β*	Part	*B*	*β*	Part	*B*	*β*	Part	*B*	*β*	Part
Letter names	0.86**	0.63**	.42	0.80**	0.61**	.40	0.73**	0.53**	.38	1.01**	0.78**	.64
Letter sounds	0.23	0.13	.09	0.31*	0.19*	.12	0.34*	0.22*	.15	-0.14	-0.09	-.07
Phoneme segmenting	0.13	0.08	.07	-0.20	-0.10	-.09	-0.06	0.13	-.03	-0.15	-0.07	-.06
R^2		.56			.54			.49			.53	
N		170			175			171			184	

Note. **p* < .05, ***p* < .001, Part = Semipartial Correlation.

Reference

Wray, K. A., Lai, C.-F., Sáez, L., Alonzo, J., & Tindal, G. (2013). *easyCBM® kindergarten beginning reading measures: Alternate form reliability and criterion validity with the SAT-10 (Technical Report #1403).* Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 2401: Criterion validity and classification accuracy of easyCBM® Grades 3-8 (Tindal & Nese, 2024).

Methods

TR2401 examines criterion validity and classification accuracy for easyCBM® reading benchmarks in Grades 3–8, using the Smarter Balanced (SB) English Language Arts (ELA) test as the external criterion. The sample includes well over 8,000 anonymized student records from two states and four school districts. Students completed easyCBM® benchmarks in fall 2023, winter 2024, and spring 2024 on three reading measures: Proficient Reading (20 items), Vocabulary (20 items), and Oral Reading Fluency (ORF; words correct per minute). SB ELA was administered in spring.

For predictive/concurrent validity, the report presents grade-specific correlation matrices (Tables 19–24 of the Technical Report) linking each easyCBM® reading measure to SB ELA in fall, winter, and spring. Because the reading measures are on different scales, the authors also compute a reading composite by converting each reading score to a grade-specific z score and averaging within grade; the composite is then correlated with SB ELA by season.

For classification accuracy, SB ELA serves as the criterion and "risk" is primarily defined at the 20th percentile. For each season and reading measure, the report provides (a) an evidence table with the SB cut score, the corresponding easyCBM® cut score, confusion-matrix counts (true/false positives/negatives), and AUC with a 95% confidence interval; and (b) a summary table reporting sample size, base rate, overall classification rate, sensitivity, specificity, false positive/negative rates, and positive/negative predictive power.

Results

Criterion validity evidence indicates that all three easyCBM® reading measures relate meaningfully to SB ELA, with strength varying by grade, season, and measure. For Proficient Reading, correlations with SB ELA range from .53 (Grade 8 fall) to .70 (Grade 4 fall), with most values in the high .50s to high .60s across grades and seasons (e.g., Grade 3: .63 fall, .60 winter, .69 spring; Grade 5: .62–.66). Vocabulary shows similar or slightly stronger relations in several grades, ranging from .55 (Grade 7 winter) to .69 (Grade 3 winter), with many values around .61–.67 (e.g., Grade 4: .67 fall/winter/spring; Grade 6: .62 fall, .58 winter, .66 spring). ORF correlations are more variable and generally lower in upper grades, ranging from .48 (Grade 6 fall) and .49 (Grade 7 winter) to .66 (Grade 3 winter) and .65 (Grade 3 fall/spring). These patterns support the interpretation that easyCBM® reading benchmarks provide moderate-to-strong predictive/concurrent evidence with SB ELA, and that alignment between the easyCBM® measure and the SB construct is strongest for Proficient Reading and Vocabulary in most grades.

The reading composite reinforces the overall relationship with SB ELA: the composite correlation is .64 in fall, .57 in winter, and .66 in spring. Taken together, these coefficients show that easyCBM® reading performance months before SB (fall/winter) is meaningfully associated with end-of-year SB ELA.

Classification accuracy results (risk defined at the SB 20th percentile) show that the reading measures can screen for risk with generally strong AUCs and favorable operating characteristics. For Proficient Reading, AUCs are strong across seasons (Fall: .80–.85; Winter: .76–.85; Spring: .80–.84). Corresponding sensitivity and specificity are typically in the .70–.80+ range (e.g., fall sensitivity .68–.83; spring specificity up to .88). Base rates cluster around .15–.24. Positive predictive power is consistently very high (often .91–.96), while negative predictive power is lower (often ~.33–.56), reflecting base rates and the sensitivity/specificity balance.

ORF also shows strong screening accuracy. AUCs are .75–.85 in fall, .77–.86 in winter, and .80–.84 in spring. Sensitivity is often strong in Grades 3–6 (e.g., winter Grade 3 = .83; spring Grade 4 = .82) but weaker in some upper-grade cells (e.g., fall Grade 7 = .62). Specificity ranges from about .59 to .85 depending on grade and season. Vocabulary yields similarly strong AUCs (Fall: .81–.84; Winter: .80–.89; Spring: .82–.87) with generally balanced sensitivity and specificity and high positive predictive power. Across measures, TR2401 concludes that easyCBM® reading benchmarks provide credible criterion validity with SB ELA and strong classification accuracy for identifying students at risk when using season- and grade-specific cut scores.

NCII-aligned summaries show overall classification rates typically around .73–.83 for Proficient Reading (e.g., fall Grade 6 = .80; spring Grade 5 = .83), about .70–.81 for ORF (with a lower value in fall Grade 7 = .65), and about .73–.83 for Vocabulary. Thus, roughly three-quarters of students are correctly classified in most grade/season combinations relative to the SB 20th-percentile criterion.

Reference

Tindal, G. & Nese, J. F. T., & (2024). *Criterion validity and classification accuracy of easyCBM®: Grades 3-8. (Technical Report # 2401).* Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 0910: Criterion-related Evidence Using easyCBM® Reading Measures and Student Demographics to Predict State Test Performance in Grades 3-8 (Tindal et al., 2009a).

Methods

This technical report examines the relationship between easyCBM® reading benchmark assessments, student demographic characteristics, and performance on the Oregon Assessment of Knowledge and Skills (OAKS) reading test. The study was designed to evaluate how well easyCBM® measures and demographic variables predict student performance on a statewide standardized reading assessment.

The study used student data from two Oregon school districts and included students in grades 3 through 8. Sample sizes varied by grade and district but generally ranged from several hundred to over one thousand students per grade. The second district included a higher proportion of economically disadvantaged students and students receiving special education services.

Three easyCBM® reading benchmark measures served as predictors: passage reading fluency, vocabulary, and multiple-choice reading comprehension. Passage reading fluency was measured as the number of words read correctly per minute during individually administered reading tasks. Vocabulary and comprehension measures were computer-based, group-administered assessments.

The outcome variable was performance on the OAKS reading test, a computer-adaptive statewide assessment with vertically scaled scores. Students completed the easyCBM® benchmarks during fall, winter, and spring testing periods. Demographic variables included gender, ethnicity, economic disadvantage, Title I status, and special education status. Data from the easyCBM® database were merged with district demographic and testing records. Analyses included correlation analyses and multiple regression models to determine how well the benchmark measures and demographic factors predicted state reading test scores.

Results

Results showed consistently strong relationships between easyCBM® reading measures and the OAKS reading test across grades and benchmark periods. Correlations between the benchmark assessments and the state test were typically above .60. Regression analyses indicated that the combined predictors explained moderate to substantial portions of variance in state reading scores, with R^2 values generally ranging from about .50 to .65. The easyCBM® reading measures consistently accounted for the largest share of explained variance. Demographic variables such as economic disadvantage and special education status were sometimes significant predictors but showed smaller and less consistent effects. Overall, the findings indicate that easyCBM® reading benchmarks provide meaningful predictive evidence for statewide reading assessment performance

Table 14. Illustrative Table from Technical Report 0910

Correlations

		Passage Reading Fluency score, fall	Vocabulary score, fall	Multiple Choice Reading Comprehension score, fall	OAKS reading score, 2008-2009
Passage Reading Fluency	Pearson Correlation	1	.715**	.629**	.628**
	Sig. (2-tailed)		.000	.000	.000
	N	338	314	315	338
Vocabulary score	Pearson Correlation	.715**	1	.640**	.676**
	Sig. (2-tailed)	.000		.000	.000
	N	314	314	314	314
Multiple Choice Reading Comprehension	Pearson Correlation	.629**	.640**	1	.624**
	Sig. (2-tailed)	.000	.000		.000
	N	315	314	315	315
OAKS reading score, 2008-2009	Pearson Correlation	.628**	.676**	.624**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	338	314	315	797

**. Correlation is significant at the 0.01 level (2-tailed).

Reference

Tindal, G., Nese, J. F., & Alonzo, J. (2009). *Criterion-related evidence using easyCBM® reading measures and student demographics to predict state test performance in Grades 3-8 (Technical Report # 0910)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1106: Diagnostic Efficiency of easyCBM® Reading: Oregon (Park, Anderson, et al., 2011).

Methods

This report examined the diagnostic efficiency of easyCBM® reading benchmark assessments for predicting performance on the Oregon Assessment of Knowledge and Skills (OAKS). Data were collected during the 2009–2010 academic year from three Oregon districts. The sample included approximately 2,000 students per grade in grades 3–8.

Measures included Passage Reading Fluency (PRF), Multiple-Choice Reading Comprehension (MCRC), and Vocabulary (VOC). Fluency and comprehension were administered fall, winter, and spring; vocabulary was administered fall and spring.

Receiver Operating Characteristic (ROC) analyses were conducted to evaluate classification accuracy. Sensitivity, specificity, and area under the curve (AUC) statistics were calculated for each measure and benchmark occasion. Cut score selection emphasized maximizing sensitivity while maintaining specificity above .70, consistent with RTI priorities aimed at minimizing false negatives. Demographic subgroup comparisons were conducted to ensure comparability across participating districts.

Results

Across grades and seasons, AUC values generally fell within acceptable to strong ranges, indicating meaningful discrimination between students who passed and did not pass OAKS. PRF demonstrated particularly strong diagnostic utility in winter and spring administrations. MCRC and VOC also demonstrated adequate discrimination, with comprehension measures often showing stronger predictive utility at upper grades.

Sensitivity was consistently prioritized in cut score selection, often exceeding .80 while maintaining acceptable specificity. Results indicated that easyCBM® benchmark measures provided strong criterion-related classification evidence when aligned with Oregon's state proficiency standards.
However, performance varied slightly by grade and season, reinforcing the importance of grade-specific cut score interpretation. Overall findings support the use of easyCBM® reading benchmarks as effective RTI screening tools in Oregon.

Table 15. Illustrative Table from Technical Report 1106

Table 30

Resulting Statistics for Each Chosen Cut Score: ELL, PRF

Grd	Season	Meets Score	*n*	Failure Base Rate	False Positive Rate	False Negative Rate	Sensitivity	Specificity	Positive Predictive Power	Negative Predictive Power	Overall Correct Classification	AUC
	Fall	62	97	0.25	0.32	0.17	0.83	0.68	0.47	0.93	0.72	0.81
3	Winter	90	97	0.26	0.39	0.20	0.80	0.61	0.42	0.90	0.66	0.82
	Spring	93	93	0.22	0.32	0.25	0.75	0.68	0.39	0.91	0.70	0.76
	Fall	96	100	0.39	0.33	0.18	0.82	0.67	0.62	0.85	0.73	0.82
4	Winter	111	77	0.38	0.25	0.14	0.86	0.75	0.68	0.90	0.79	0.87
	Spring	126	81	0.37	0.37	0.10	0.90	0.63	0.59	0.91	0.73	0.88
	Fall	138	95	0.49	0.54	0.11	0.89	0.46	0.62	0.81	0.67	0.75
5	Winter	142	75	0.44	0.50	0.15	0.85	0.50	0.57	0.81	0.65	0.76
	Spring	160	90	0.46	0.49	0.17	0.83	0.51	0.59	0.78	0.66	0.79
	Fall	135	48	-	-	-	-	-	-	-	-	-
6	Winter	149	25	-	-	-	-	-	-	-	-	-
	Spring	154	35	-	-	-	-	-	-	-	-	-
	Fall	144	145	0.77	0.76	0.12	0.88	0.24	0.79	0.38	0.73	0.66
7	Winter	160	140	0.76	0.79	0.08	0.92	0.21	0.79	0.44	0.75	0.75
	Spring	152	154	0.77	0.67	0.12	0.88	0.33	0.81	0.46	0.75	0.69
	Fall	166	149	0.85	0.70	0.06	0.94	0.30	0.88	0.50	0.85	0.78
8	Winter	157	150	0.85	0.57	0.07	0.93	0.43	0.90	0.53	0.85	0.80
	Spring	160	156	0.83	0.50	0.05	0.95	0.50	0.90	0.65	0.87	0.81

Note. AUC = Area Under the ROC Curve

Reference

Park, B. J., Anderson, D., Irvin, P. S., Alonzo, J., & Tindal, G. (2011). *Diagnostic efficiency of easyCBM® reading: Oregon (Technical Report No. 1106)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1107: Diagnostic Efficiency of easyCBM® Reading: Washington State (Anderson, Park, et al., 2011).

Methods

This study evaluated the diagnostic efficiency of easyCBM® reading benchmark assessments in predicting performance on Washington's Measurements of Student Progress (MSP). Data were collected during the 2009–2010 academic year from two Washington districts. The sample included approximately 900 students per grade in grades 3–8. Measures included PRF, MCRC, and VOC administered across fall, winter, and spring benchmark windows. ROC curve analyses were conducted for each measure and grade. AUC statistics were calculated to evaluate overall discrimination, and sensitivity and specificity were examined for all potential cut scores. Cut score decisions emphasized minimizing false negatives, consistent with RTI implementation priorities. Demographic comparisons confirmed that subgroup representation did not meaningfully bias classification outcomes.

Results

AUC statistics demonstrated acceptable to strong classification accuracy across grades and measures. PRF generally yielded strong discrimination, particularly in winter and spring. MCRC and VOC demonstrated moderate to strong diagnostic utility depending on grade level. Sensitivity values were typically emphasized above specificity in final cut score decisions, reflecting the goal of identifying students at risk of failing MSP. Classification accuracy was strongest in lower and middle grades and showed some variability in upper grades. Overall, findings support the use of easyCBM® reading benchmarks for predicting MSP outcomes and informing RTI tier placement decisions in Washington.

Table 16. Illustrative Table from Technical Report 1107

Table 5

Resulting Statistics for Each Chosen Cut Score: Full Sample, PRF

Grd	Season	Meets Score	*n*	Failure Base Rate	False Positive Rate	False Negative Rate	Sensitivity	Specificity	Positive Predictive Power	Negative Predictive Power	Overall Correct Classification	AUC
3	Fall	79	253	0.36	0.29	0.15	0.85	0.71	0.62	0.89	0.76	0.84
	Winter	106	285	0.35	0.29	0.22	0.78	0.71	0.59	0.86	0.73	0.81
	Spring	101	310	0.35	0.30	0.21	0.79	0.70	0.59	0.86	0.73	0.82
4	Fall	109	131	0.44	0.49	0.26	0.74	0.51	0.54	0.71	0.61	0.66
	Winter	124	245	0.38	0.38	0.26	0.74	0.62	0.55	0.79	0.67	0.71
	Spring	132	259	0.40	0.40	0.24	0.76	0.60	0.56	0.79	0.66	0.76
5	Fall	130	146	0.42	0.30	0.18	0.82	0.70	0.67	0.84	0.75	0.84
	Winter	140	248	0.44	0.30	0.29	0.71	0.70	0.66	0.75	0.71	0.80
	Spring	161	228	0.43	0.40	0.19	0.81	0.60	0.60	0.81	0.69	0.78
6	Fall	145	231	0.44	0.39	0.30	0.70	0.61	0.59	0.72	0.65	0.74
	Winter	176	241	0.45	0.38	0.25	0.75	0.62	0.62	0.75	0.68	0.77
	Spring	177	61	0.57	0.38	0.29	0.71	0.62	0.71	0.62	0.67	0.75
7	Fall	143	198	0.36	0.30	0.24	0.76	0.70	0.59	0.84	0.72	0.78
	Winter	165	428	0.41	0.30	0.30	0.70	0.70	0.62	0.77	0.70	0.76
	Spring	152	259	0.38	0.40	0.29	0.71	0.60	0.52	0.78	0.64	0.72
8	Fall	144	486	0.39	0.30	0.22	0.78	0.70	0.62	0.83	0.73	0.82
	Winter	153	508	0.37	0.29	0.26	0.74	0.71	0.60	0.82	0.72	0.81
	Spring	174	510	0.36	0.30	0.19	0.81	0.70	0.61	0.86	0.74	0.80

Note. AUC = Area Under the ROC Curve

Reference

Anderson, D., Park, B. J., Irvin, P. S., Alonzo, J., & Tindal, G. (2011). *Diagnostic efficiency of easyCBM® reading: Washington State (Technical Report No. 1107)*. University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1108: Cross-validation of easyCBM® Reading Cut Scores in Oregon: 2009–2010 (Grades 3 – 8) (Park, Irvin, et al., 2011).

Methods

This report extended prior diagnostic efficiency work by conducting a cross-validation study of easyCBM® reading benchmark cut scores in Oregon. Data from approximately 2,000 students per grade (3–8) were randomly split into two demographically comparable groups. Measures included PRF, MCRC, and VOC administered across seasonal benchmarks. ROC analyses were conducted separately for each split sample. AUC statistics and sensitivity/specificity values were computed for all possible cut scores. Cut score selection followed modified Silberglitt and Hintze decision rules, emphasizing sensitivity thresholds above .70 and preferably above .80 while maintaining acceptable specificity. Stability of optimal cut scores was evaluated by comparing differences between groups and examining overlap of 95% confidence intervals for AUC statistics.

Results

Optimal cut scores demonstrated strong stability across randomly split groups. The average difference between PRF cut scores across groups was approximately 1–2 CWPM across grade-level comparisons. Differences for MCRC and VOC were typically less than one raw score point. AUC confidence intervals overlapped across groups for all grades and benchmark occasions, indicating non-significant differences in classification accuracy. Sensitivity and specificity patterns remained consistent across groups. These findings provide strong cross-validation evidence that previously identified Oregon cut scores are statistically stable and robust. However, authors caution that cut scores are criterion-dependent and should not be generalized across states without validation.

Table 17. Illustrative Table from Technical Report 1108

Grade 8
Spring PRF Benchmark

Cut score	Group 1		Group 2	
	Sensitivity	Specificity	Sensitivity	Specificity
14.00	.000	1.000	-	-
17.00	.002	1.000	-	-
21.00	-	-	.000	1.000
22.50	.005	1.000	-	-
25.00	-	-	.002	1.000
30.50	-	-	.005	1.000
31.50	.007	1.000	-	-
35.50	-	-	.007	1.000
40.00	.010	1.000	-	-
41.50	-	-	.009	1.000
48.00	.012	1.000	.012	1.000
53.50	.015	1.000	-	-
54.00	-	-	.014	1.000
54.50	.017	1.000	-	-
57.00	.020	1.000	-	-
57.50	-	-	.016	1.000
59.00	-	-	.019	1.000
60.50	.025	1.000	-	-
61.00	-	-	.021	1.000
62.50	.027	1.000	.026	1.000
63.50	.029	1.000	.028	1.000
64.50	-	-	.031	1.000
65.50	-	-	.035	1.000
66.00	.032	1.000	-	-
67.50	-	-	.040	1.000
69.50	-	-	.042	1.000
70.50	.034	1.000	.045	1.000
71.50	-	-	.047	.999
72.50	-	-	.049	.999
73.50	-	-	.052	.999
74.00	.037	1.000	-	-
74.50	-	-	.056	.999
75.50	.039	1.000	.061	.999
77.00	-	-	.066	.999
77.50	.042	1.000	-	-
78.50	-	-	.071	.999
79.50	.044	1.000	.073	.999

Reference

Park, B. J., Irvin, P. S., Anderson, D., Alonzo, J., & Tindal, G. (2011). *Cross-validation of easyCBM® reading cut scores in Oregon: 2009–2010 (Technical Report No. 1108)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1109: Cross-validation of easyCBM® Reading Cut Scores in Washington: 2009–2010 (Irvin et al., 2011).

Methods

This report conducted a cross-validation study of easyCBM® reading benchmark cut scores aligned with Washington's MSP assessment. Approximately 900 students per grade (3–8) were randomly divided into two comparable samples. Measures included PRF, MCRC, and VOC administered across fall, winter, and spring. ROC analyses were conducted separately within each group. AUC statistics were compared using overlapping 95% confidence intervals. Cut scores were selected using sensitivity-prioritized decision rules designed to minimize false negatives in RTI implementation. Demographic comparability between groups was examined via t-tests across ethnicity, SPED, ELL, and FRL categories.

Results

Optimal cut scores were reasonably stable across randomly split groups, though slightly more variable than observed in Oregon. Average PRF cut score differences were approximately 7 CWPM across grade-level comparisons. Differences for MCRC and VOC were generally under one point. AUC confidence intervals overlapped for all measures and grades, indicating statistically equivalent classification accuracy across groups. Sensitivity values were typically maintained at or above .80 while specificity remained acceptable. Overall, findings provide strong evidence that identified Washington cut scores are stable and defensible for MSP classification prediction. Authors note that cut scores should be interpreted within state-specific accountability contexts.

Table 18. Illustrative Table from Technical Report 1109

Grade 4
Fall MCRC Benchmark

Cut score	Group 1		Group 2	
	Sensitivity	Specificity	Sensitivity	Specificity
-1.00	.000	1.000	.000	1.000
1.50	-	-	.000	.973
2.00	.037	.972	-	-
3.50	-	-	.032	.973
4.50	.111	.972	.065	.946
5.50	.185	.972	.097	.919
6.50	.259	.917	.097	.865
7.50	.370	.889	.355	.838
8.50	.444	.806	.516	.784
9.50	.630	.778	**.710**	**.757**
10.50	.667	.722	.774	.649
11.50	**.741**	**.639**	.871	.622
12.50	.815	.528	.935	.486
13.50	.852	.361	.968	.297
14.50	.926	.361	1.000	.297
15.50	1.000	.250	1.000	.216
16.50	1.000	.167	1.000	.108
17.50	1.000	.139	1.000	.081
18.50	1.000	.028	1.000	.027
20.00	1.000	.000	1.000	.000

Reference

Irvin, P. S., Park, B. J., Anderson, D., Alonzo, J., & Tindal, G. (2011). *Cross-validation of easyCBM® reading cut scores in Washington: 2009–2010 (Technical Report No. 1109)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1302: Examining the Internal Structure of the easyCBM® Reading Measures, K–5 (Alonzo et al., 2013c).

Methods

This technical report investigated the **internal structure** (construct validity evidence) of the easyCBM® CCSS reading measures for Grades K–5. The study used student assessment data collected during a defined administration window and evaluated whether the measures' observed score patterns matched the hypothesized constructs of early literacy, fluency, and comprehension across grade levels. The authors applied confirmatory factor analysis (CFA) to test competing measurement models, including one-factor versus multi-factor structures and versions that accounted for "testlet effects" (shared variance among grouped items, such as comprehension questions tied to the same passage). Analyses were organized by grade bands (notably K–2 versus 3–5) to reflect developmental differences in reading skills and the structure of the assessment system. The report examined parameter estimates such as factor loadings (how strongly each measure related to its intended construct), correlations among latent constructs, and variance explained in observed indicators. Attention was given to data features that can distort CFA results, including non-normal score distributions and ceiling effects on some components, and the authors interpreted models with those constraints in mind. Fit was evaluated using standard indices and comparative evidence across models rather than relying on a single statistic.

Therefore, this report also provides important construct validity evidence for easyCBM® reading measures using confirmatory factor analysis (CFA). Students in grades K–5 completed a battery of easyCBM®

reading assessments. For grades K–2, measures reflect early literacy (Phoneme Segmenting, Letter Names, Letter Sounds), fluency (two Word Reading Fluency forms and Passage Reading Fluency), and comprehension (CCSS-style comprehension and MCRC). Competing CFA models (e.g., one-factor vs multi-factor; with/without testlet effects) are compared using model fit indices, and standardized factor loadings are examined to evaluate alignment of measures with hypothesized constructs.

Results

Findings generally supported a multi-component reading structure that aligns with how the easyCBM® CCSS system is intended to function, while also flagging some grade-band-specific complications. For Grades K–2, the early literacy indicators (e.g., letter knowledge and phonological measures) and fluency indicators tended to show strong relationships to their intended factors, while some comprehension components demonstrated weaker associations under certain model specifications. Where comprehension performance showed heavy skew (including strong ceiling effects), model estimates for comprehension paths were less stable and appeared to influence overall fit and covariance patterns. When alternative model specifications were examined, the comprehension indicators typically showed improved associations with the comprehension construct, and the overall model became more interpretable. Across analyses, the latent constructs showed meaningful relationships with one another (generally moderate associations), consistent with the idea that early literacy, fluency, and comprehension are related but distinguishable skills. Overall, the results provided supportive evidence for the construct validity of the K–5 easyCBM® CCSS reading measures, while also identifying places where components (especially early comprehension subparts in lower grades) may require careful interpretation due to distributional issues.

For the K–2 sample, CFA results support a multi-dimensional reading structure and report that a three-factor model (Early Literacy, Fluency, Comprehension) provides the best fit among the tested models. Standardized factor loadings (lambda coefficients) are strong for most early literacy and fluency indicators: Phoneme Segmenting loads moderately (.56), Letter Names strongly (.89), and Letter Sounds strongly (~.82) on the Early Literacy factor; fluency indicators load strongly (e.g., WRF1 .88–.89; WRF2 .80–.81; PRF .84–.85). Comprehension indicators vary: one CCSS comprehension subcomponent shows a near-zero loading (.03), consistent with weak association/ceiling effects, whereas MCRC comprehension shows moderate loadings (.55–.67). Factor correlations show Early Literacy and Fluency are moderately related (.61–.62), while Fluency and Comprehension show a moderate negative relation in one model (about −.33), and some early literacy–comprehension relations are small/non-significant in the reported solutions. Overall, the findings support interpreting easyCBM® reading measures as assessing theoretically coherent but distinct constructs at K–2, strengthening the construct validity argument for these measures.

Table 19. Example Construct Validity Output

Table 3

CFA Model Fit Summary for Grade K-2 Sample

	Grades K-2			
	1st Model	2nd Model	**3rd Model**	4th Model
CFI	0.56	0.54	**0.77**	0.73
TLI	0.38	0.41	**0.62**	0.62
RMSEA	0.31	0.25	**0.24**	0.20
SRMR	0.15	0.13	**0.08**	0.08
AIC	8254.31	10764.87	**8136.82**	10650.39
BIC	8322.76	10850.43	**8213.82**	10744.50
ABIC	8246.86	10755.56	**8128.44**	10640.14

1st Model: One factor CFA model without testlet effect
2nd Model: One factor CFA model with testlet effect
3rd Model: Three factor CFA model without testlet effect
4th Model: Three factor CFA model with testlet effect

Table 4
Maximum Likelihood Parameter Estimates of Three Factor Confirmatory Factor Analysis Model with Testlet Effects for Grades K-2

Parameters	Early Literacy 3rd	Early Literacy 4th	Fluency 3rd	Fluency 4th	Comprehension 3rd	Comprehension 4th
Lambda coefficients						
Phoneme segmenting	0.56	0.54				
Letter names	0.89	0.89				
Letter sounds	0.82	0.82				
WRF1			0.88	0.89		
WRF2			0.81	0.80		
PRF			0.85	0.84		
Comprehension → Part1A						0.73
Comprehension → Part1B					0.78	0.61
Comprehension → Part1C						0.03
Comprehension → Part2					0.67	0.55
Factor correlation						
		Model 3			Model 4	
Early Literacy ↔ Fluency		0.61			0.62	
Early Literacy ↔ Comprehension		0.11 NS			-0.12NS	
Fluency ↔ Comprehension		-0.33			-0.54	
Variance explained by the model						
		Model 3			Model 4	
PS		0.31			0.29	
LN		0.79			0.79	
LS		0.67			0.68	
WRF1		0.77			0.79	
WRF2		0.66			0.64	
PRF		0.71			0.70	
Part1A					0.53	
Part1B		0.60			0.38	
Part1C					<0.01	
Part2		0.45			0.30	

Note. Parameter estimates reported in this table are standardized coefficients.

Table 20. Illustrative Table from Technical Report 1302

Table 12
Maximum Likelihood Parameter Estimates of Two Factor Confirmatory Factor Analysis Model with Testlet Effects for Grades 3-5

Parameter	Grade 3 Fluency	Grade 3 Comp	Grade 4 Fluency	Grade 4 Comp	Grade 5 Fluency	Grade 5 Comp
Lambda coefficients						
Fluency → WRF	0.82		0.91		0.90	
Fluency → PRF1	0.97		0.97		0.99	
Fluency → PRF2	0.95		0.98		0.96	
Comprehension → Part1A		0.42		0.44		0.96
Comprehension → Part1B		0.88		0.20		0.97
Comprehension → Part1C		0.74		0.72		0.70
Comprehension → Part2		0.34		0.81		-0.37
Factor correlation						
Fluency ↔ Comprehension	0.47		0.04NS		-0.56	
Variance explained by the model						
WRF	0.64		0.83		0.81	
PRF1	0.95		0.94		0.99	
PRF2	0.91		0.95		0.92	
Part1A	0.18		0.20		0.91	
Part1B	0.78		0.04		0.94	
Part1C	0.55		0.51		0.49	
Part2	0.12		0.66		0.14	

Note. Parameter estimates reported in this table are standardized coefficients.

Reference

Alonzo, J., Park, B. J., & Tindal, G. (2013). *Examining the internal structure of the easyCBM® reading measures, grades K–5* (Technical Report #1302). Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1304: An Examination of the Internal Structures of the Gr. 6–8 easyCBM CCSS Reading Measures: A Construct Validity Study (Alonzo et al., 2013a).

Methods

This study examined the internal structure of the Grades 6–8 easyCBM® CCSS Reading assessments using confirmatory factor analysis. The purpose was to determine whether the assessments functioned as unidimensional measures of literal reading comprehension or whether distinct factors emerged based on passage genre. Participants included 482 students in Grades 3–5 drawn from multiple states; the design was later extended to middle grades. Each test form contained 45 items organized into nine testlets representing three genres: Read to Perform a Task, Informational Text, and Short Literary Text. Each testlet included five selected-response items. Using maximum likelihood estimation, researchers tested one-factor, two-factor, and three-factor models. Model fit indices including CFI, TLI, RMSEA, and chi-square statistics were evaluated. Factor loadings, variance estimates, and squared multiple correlations were examined to determine dimensional coherence.

Results

Across grades, results consistently supported a unidimensional interpretation of the CCSS Reading measures. Although certain two-factor models demonstrated comparable fit, factor correlations were high, indicating that genre distinctions did not meaningfully separate constructs. Three-factor models were not supported. The one-factor model provided acceptable fit across grades, supporting the interpretation that the assessments measure a broad construct of literal reading comprehension. The findings reinforce the design intention that items drawn from different genres contribute to a unified comprehension construct rather than distinct subskills.

Table 21. Illustrative Table from Technical Report 1304

Table 6
Grade 3 Regression Weights, Two-Factor Model (Read to Perform a Task & Informational Text/Short Literary Text)

	Unstandardized				Standardized
	Estimate	S.E.	C.R.	P	
T1 <--- F1	1.00				.69
T2 <--- F1	0.73	.12	6.35	***	.64
T3 <--- F1	1.12	.14	8.13	***	.83
T4 <--- F2	1.00				.83
T5 <--- F2	1.06	.07	11.09	***	.85
T6 <--- F2	1.03	.09	11.31	***	.87
T7 <--- F2	1.00	.10	10.15	***	.81
T8 <--- F2	0.98	.11	9.06	***	.75
T9 <--- F2	0.97	.12	8.04	***	.67

Reference

Alonzo, J., Park, B. J., & Tindal, G. (2013). *An examination of the internal structures of the easyCBM CCSS reading measures (Technical Report # 1304).* Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1305: An Examination of the Internal Structures of the Gr. K–5 easyCBM® CCSS Reading Measures: A Construct Validity Study (Alonzo et al., 2013b).

This study examined the internal structure and construct validity of the easyCBM® K–5 reading measures using confirmatory factor analysis (CFA). The purpose was to determine whether the assessments measured a single general reading construct or multiple related but distinct constructs, and whether genre or measure type altered the factor structure.

Methods

The study was conducted in winter 2012 with 22 teachers and 519 students across kindergarten through Grade 5. Participants were recruited from multiple states and included students with disabilities and English learners. Teachers administered the assessments using standardized protocols.

Different measures were administered by grade band. In Grades K–2, students completed early literacy measures (phoneme segmenting, letter names, letter sounds), word reading fluency, passage reading fluency, and two comprehension measures (CCSS and MCRC). Fluency measures were individually administered, while comprehension measures were group administered. In Grades 3–5, students completed word reading fluency, two passage reading fluency measures, and both CCSS and MCRC comprehension assessments.

Four CFA models were tested for each grade band using maximum likelihood estimation in Mplus. For Grades K–2, models compared one-factor versus three-factor solutions, with and without testlet effects. For Grades 3–5, models compared one-factor and two-factor solutions, again with and without testlet effects. Model fit was evaluated using CFI, TLI, RMSEA, SRMR, and information criteria (AIC, BIC, ABIC). Due to negative skew in CCSS comprehension scores, collapsed scoring procedures were implemented to address convergence issues.

Results

For Grades 3–5, results varied somewhat by grade. Both one-factor and two-factor models showed acceptable fit, but the two-factor model separating fluency and comprehension generally demonstrated slightly better theoretical and statistical support. Fluency indicators consistently showed strong loadings. Comprehension measures also loaded meaningfully, despite differences in passage length and genre between CCSS and MCRC measures. Importantly, both comprehension types loaded on a single comprehension factor, suggesting that shorter, more accessible CCSS measures assess the same underlying construct as longer MCRC measures.

Overall, the findings support the construct validity of the easyCBM® reading measures. In early grades, reading appears best represented by multiple interrelated constructs, while in upper elementary grades, fluency and comprehension remain distinguishable but more closely integrated. The results provide evidence that the system's measures align with their intended theoretical constructs and can be used confidently for screening and progress monitoring purposes.

Table 22. Illustrative Table from Technical Report 1305

Table 11
CFA Model Fit Summary for Grades 3-5

Grade 3					Grade 4				Grade 5			
	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th
CFI	0.96	0.83	-	**0.84**	0.94	0.86	-	**0.98**	0.97	0.53	1.00	**0.96**
TLI	0.93	0.74	-	**0.74**	0.88	0.80	-	**0.97**	0.95	0.29	1.01	**0.94**
RMSEA	0.17	0.24	-	**0.24**	0.23	0.21	-	**0.08**	0.16	0.49	0.00	**0.15**
SRMR	0.06	0.12	-	**0.15**	0.11	0.14	-	**0.07**	0.07	0.18	0.01	**0.09**
AIC	3529.52	4191.76	-	**4187.88**	3983.52	4263.06	-	**4207.04**	2485.10	3144.17	2475.93	**2933.84**
BIC	3568.45	4246.26	-	**4244.97**	4023.61	4319.19	-	**4265.84**	2518.39	3190.78	2511.44	**2982.67**
ABIC	3521.08	4179.94	-	**4175.49**	3976.22	4252.84	-	**4196.33**	2471.16	3124.65	2461.06	**2913.39**

1st Model: One-factor CFA model without testlet effect
2nd Model: One-factor CFA model with testlet effect
3rd Model: Two-factor CFA model without testlet effect
4th Model: Two-factor CFA model with testlet effect
* - Model did not converge or indicated significant errors, thus the results are not interpreted

Reference

Alonzo, J., Park, B. J., & Tindal, G. (2013). *An examination of the internal structures of the Gr. K–5 easyCBM CCSS reading measures: A construct validity study (Technical Report No. 1305).* Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 0025: Analysis of Reading Fluency and Comprehension Measures for First Grade Students (Alonzo & Tindal, 2004a).

This technical report (Technical Report 0025) presents a comprehensive evaluation of reading assessment measures designed for universal screening, benchmarking, and progress monitoring.

Methods

The study draws upon large and diverse student samples representing multiple grade levels. Standardized administration protocols were implemented to ensure consistency across testing sites. Analyses included classical test theory indices such as internal consistency reliability, item-total correlations, and standard error of measurement. Item response theory modeling was applied where appropriate to examine item difficulty calibration, discrimination parameters, and overall model fit. Dimensionality analyses were conducted to evaluate whether the measures captured a unidimensional construct of reading proficiency or reflected multiple latent traits. Additional procedures included scaling analyses to ensure vertical alignment across grade levels, subgroup analyses to explore differential item functioning, and growth modeling to assess sensitivity to instructional change. Classification accuracy statistics were calculated to evaluate screening cut scores and decision consistency. Together, these analytic procedures were designed to establish both technical adequacy and practical utility for educational decision-making.

Results

Results indicate that the reading measures demonstrate acceptable to strong reliability across grades and administration periods. Reliability coefficients consistently met or exceeded commonly accepted benchmarks for screening and progress monitoring applications. Item analyses revealed that most items functioned within recommended statistical thresholds for difficulty and discrimination, with only a small number flagged for further review due to misfit or instability. Dimensionality findings generally supported a dominant reading proficiency construct, while scaling analyses confirmed appropriate score interpretation across grade levels. Classification accuracy indices demonstrated adequate sensitivity and specificity for identifying students at risk. **Predictive and concurrent validity** evidence showed meaningful relationships between these measures and external indicators of reading performance. Growth analyses suggested that the assessments were sensitive to developmental change over time, supporting their use for longitudinal progress monitoring. Overall, findings support the technical soundness of these reading assessment measures and their suitability for informing instructional planning, intervention decisions, and system-level educational evaluation.

Reference

Alonzo, J., & Tindal, G. (2004). *Analysis of reading fluency and comprehension measures for first grade students (Technical Report No. 0025).* Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 0022: Analysis of Reading Fluency and Comprehension Measures (3rd Grade) (Ketterlin-Geller & Tindal, 2004).

Methods

The study analyzed third-grade student data collected during a district-wide reading assessment administration. Measures included timed oral reading fluency passages and structured comprehension probes aligned with grade-level standards. Oral reading fluency was operationalized as words read correctly per minute. Comprehension performance was evaluated using passage-based response items. Statistical analyses included correlation analysis between fluency and comprehension scores, internal consistency reliability estimates, and score distribution evaluations. Subgroup analyses examined patterns across achievement levels to determine the extent to which fluency functioned as a proxy for comprehension performance.

Results

Results indicated a strong relationship between oral reading fluency and comprehension performance in third grade. Students with higher fluency rates tended to demonstrate stronger comprehension outcomes. Reliability estimates for both measures were within acceptable ranges for screening and progress monitoring .Fluency measures effectively identified students at risk for reading difficulties, while comprehension measures added depth to instructional interpretation. Overall findings supported the technical adequacy of the third grade reading assessment system and confirmed fluency as a robust indicator of reading proficiency at this grade level.

Table 23. Illustrative Table from Technical Report 0022

Table 7

Regression Summary for the Composite State Score in Reading.

	Unstandardized Coefficients		Standardized Coefficients		95% Confidence Interval for B	
Independent Variables	B	Std. Error	Beta	t	Lower Bound	Upper Bound
ORF	0.09	0.02	.19	5.65*	0.06	0.13
Reading Comprehension	0.66	0.44	.04	1.48	-0.21	1.52
Vocabulary	2.18	0.22	.34	9.78*	1.75	2.62
Constant	150.93	4.56		33.07*	141.98	156.89

* $p < .01$

Reference

Ketterlin-Geller, L. R., & Tindal, G. (2004). *Analysis of reading fluency and comprehension measures for 3rd-grade students (Technical Report # 22).* Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 0027: Analysis of Reading Fluency and Comprehension Measures for Fourth Grade Students (Alonzo & Tindal, 2004b).

Methods

This report documents the development and evaluation of fourth grade reading fluency and comprehension measures intended for screening and progress monitoring. Student samples were drawn from multiple instructional contexts to ensure variability in performance. Standardized administration procedures were implemented. Passage construction and alternate-form development followed structured technical specifications to ensure comparability across administrations. Classical test theory indices were calculated to evaluate internal consistency and item discrimination. Rasch modeling procedures were used to calibrate item difficulty and examine fit statistics. Dimensionality analyses examined the coherence of reading constructs, and classification accuracy statistics were computed to evaluate screening thresholds.

Results

Results indicate acceptable to strong reliability for screening and progress monitoring. Reliability coefficients met commonly accepted benchmarks for educational decision-making. Most items demonstrated appropriate difficulty and discrimination levels. Rasch modeling supported stable calibration across forms and student samples. Dimensionality analyses supported interpretation of the measures as coherent reading constructs. Alternate forms demonstrated sufficient equivalency to support growth interpretation over time. **Classification accuracy indices showed adequate sensitivity and specificity** for identifying at-risk students. Growth analyses confirmed responsiveness to developmental change. Overall findings support the technical adequacy and instructional utility of the fourth grade reading fluency and comprehension measures.

Table 24. Illustrative Table from Technical Report 0027

Table 12

Correlations Between the Grade 4 Measures, Form A of CR

		District ORF	District Voc.	District SR Rdg	District CR Rdg	State Rdg
District ORF	Pearson Correlation	1	.363**	.444**	.332**	.605**
	Sig. (2-tailed)	.	.000	.000	.000	.000
	n	585	583	583	583	502
District Voc.	Pearson Correlation		1	.362**	.301**	.442**
	Sig. (2-tailed)		.	.000	.000	.000
	n		594	590	590	510
District SR Reading, Form A	Pearson Correlation			1	.195**	.513**
	Sig. (2-tailed)			.	.000	.000
	n			591	591	506
District CR Reading, Form A	Pearson Correlation				1	.344**
	Sig. (2-tailed)				.	.000
	n				591	506
State Reading	Pearson Correlation					1
	Sig. (2-tailed)					.
	n					511

**. Correlation is significant at the .01 level (2-tailed).

Reference

Alonzo, J., & Tindal, G. (2004). *Analysis of reading fluency and comprehension measures for fourth grade students (Technical Report No. 0027).* Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 0028: Analysis of Reading Fluency and Comprehension Measures for Fifth Grade Students (Beghetto & Tindal, 2004).

Methods

This technical report analyzes the fifth-grade District Reading Assessment, which included three primary subtests: Oral Reading Fluency (ORF), District Vocabulary, and District Reading Comprehension. The study summarized spring 2003 achievement data from 1,443 fifth-grade students across 29 schools in a mid-sized Pacific Northwest district. Three data sources were merged: District Reading Comprehension, District Vocabulary, and state reading achievement scores.

Data cleaning procedures included verification of scoring formulas, cross-checking student identifiers, and resolving duplicate records. Independent variables examined included gender, ethnicity, Special Education (SPED) status, English Language Learner (ELL) status, and school income level (coded into two regional groupings).

Form comparability analyses were conducted for ORF and Reading Comprehension alternate forms. Descriptive statistics were computed for each measure, including means and standard deviations. Item-level analyses examined item difficulty and discrimination indices. The purpose of the analysis was threefold: evaluate cross-form comparability, examine demographic performance differences, and explore abbreviation of longer forms to improve efficiency.

Results

Results indicated strong comparability between ORF alternate forms, with no statistically significant differences in form difficulty. In contrast, the Reading Comprehension forms demonstrated significant differences in difficulty between Form A and Form B, raising concerns about longitudinal interpretability across forms.

Demographic analyses revealed statistically significant differences across most subgroup comparisons. ORF scores differed significantly across gender, ethnicity, SPED, ELL status, and income level. Vocabulary results showed significant differences across all demographic variables except gender. For Reading Comprehension, Form A demonstrated significant group differences across all demographic categories except ethnicity, while Form B showed differences across all variables except school income level.
Item analyses revealed a range of difficulty levels and acceptable discrimination indices for most items. However, differences in alternate form difficulty suggested the need for further calibration to improve cross-form equivalence. Overall, the measures demonstrated adequate technical characteristics but required refinement to ensure fairness and comparability in district-level decision making.

Table 25. Illustrative Table from Technical Report 0028

Table 9
Item Analysis: CRs

Form	Item	Difficulty	Differentiation (Item-Total Correlation)
A	1	66%	.65
A	2	63%	.71
A	3	60%	.76
A	4	55%	.76
B	1	69%	.65
B	2	59%	.75
B	3	74%	.58
B	4	56%	.73

Reference

Beghetto, R., & Tindal, G. (2004). *Analysis of reading fluency and comprehension measures for fifth grade students (Technical Report No. 0028).* Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 0024: Analysis of Reading Fluency and Comprehension Measures for Sixth Grade Students (Alonzo & Tindal, 2004d).

Methods

This study examined sixth grade reading data collected during a district-wide administration of oral reading fluency and comprehension measures. Students completed timed passage readings scored as words read correctly per minute, followed by comprehension probes aligned with grade-level instructional standards. The purpose was to evaluate the technical adequacy of both measures for screening and progress monitoring in upper elementary settings. Analyses included descriptive statistics, correlation analyses between fluency and comprehension scores, and estimates of internal consistency reliability. The researchers also evaluated score distributions to identify possible ceiling or floor effects and to determine how effectively each measure differentiated among students across varying performance levels. Comparisons were conducted to determine whether fluency alone sufficiently represented overall reading proficiency at this grade level.

Results

Results indicated a meaningful but reduced relationship between oral reading fluency and comprehension compared to lower elementary grades. While students with higher fluency rates generally demonstrated stronger comprehension performance, the correlation was moderate rather than strong.

Reliability estimates for both fluency and comprehension measures were acceptable for instructional decision-making. Distributional analyses suggested that fluency measures began to show less variability among higher-performing students, whereas comprehension probes captured a broader range of performance differences.

The findings suggest that although oral reading fluency remains a useful indicator in sixth grade, comprehension measures provide additional explanatory power as texts become more complex. Together, the measures offered complementary information that strengthened the overall validity of the district's reading assessment system.

Table 26. Illustrative Table from Technical Report 0024

Table 13

Correlations Between the Grade 6 Measures

		District ORF	District Voc.	District SR Rdg	District CR Rdg	State Rdg
District ORF	Pearson Correlation	1	.29**	.47**	.40**	.54**
	Sig. (2-tailed)	.	.000	.000	.000	.000
	n	202	202	202	202	169
District Voc.	Pearson Correlation		1	.27**	.38**	.21**
	Sig. (2-tailed)		.	.000	.000	.005
	n		221	221	221	185
District SR Reading	Pearson Correlation			1	.47**	.64**
	Sig. (2-tailed)			.	.000	.000
	n			221	221	185
District CR Reading	Pearson Correlation				1	.33**
	Sig. (2-tailed)					.000
	n					185
State Reading	Pearson Correlation					1
	Sig. (2-tailed)					.
	n					185

**. Correlation is significant at the .01 level (2-tailed).

Reference

Alonzo, J., & Tindal, G. (2004). *Analysis of reading fluency and comprehension measures for sixth grade students (Technical Report No. 0024)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 0023: Analysis of Reading Fluency and Comprehension Measures for Seventh Grade Students (Alonzo & Tindal, 2004c).

Methods

The study examined seventh-grade assessment data from a district-wide administration of reading measures. Students completed timed oral reading fluency passages and grade-level comprehension assessments. Fluency was measured as words read correctly per minute, while comprehension was assessed through structured response items linked to narrative and informational texts.

Analyses included correlation estimates between fluency and comprehension, internal consistency reliability calculations, and evaluation of score distributions. Performance patterns were examined across achievement levels to assess whether fluency maintained predictive utility in middle school contexts.

Results

Findings demonstrated a moderate relationship between fluency and comprehension at the seventh-grade level. Although fluency remained associated with comprehension outcomes, the relationship was weaker than typically observed in elementary grades. Reliability estimates were acceptable for educational decision-making.

Comprehension measures captured additional variance beyond fluency, particularly among higher-performing students. The results suggest that while fluency remains useful, comprehension measures play an increasingly important role in evaluating reading competence at the secondary level.

Table 27. Illustrative Table from Technical Report 0023

Table 17

Correlation Matrix for the Different District Tests

		District ORF	District Voc.	District SR Rdg	District CR Rdg
District ORF	Pearson Correlation	1	.43**	.27**	.43**
	Sig. (2-tailed)	.	.000	.000	.000
	n	141	141	141	141
District Voc.	Pearson Correlation		1	.26**	.44**
	Sig. (2-tailed)		.	.002	.000
	n		146	146	146
District SR Reading	Pearson Correlation			1	.30**
	Sig. (2-tailed)			.	.000
	n			146	146
District CR Reading	Pearson Correlation				1
	Sig. (2-tailed)				.
	n				146

**. Correlation is significant at the .01 level (2-tailed).

Reference

Alonzo, J., & Tindal, G. (2004). *Analysis of reading fluency and comprehension measures for seventh grade students (Technical Report No. 0023).* Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1310: easyCBM® Reading Criterion Related Validity Evidence: Grades 2–5 (Lai et al., 2013a).

Methods

This technical report examined criterion-related validity evidence for the easyCBM® reading measures administered in Grades 2 through 5. The purpose of the study was to evaluate the extent to which easyCBM® vocabulary, comprehension, and fluency measures relate to established, standardized reading assessments with documented reliability and validity evidence. Participants were drawn from a convenience sample of ten schools within one Oregon district implementing easyCBM® as part of an RTI framework. Data collection occurred in January 2013 and extended to approximately six weeks due to high absenteeism. Initial recruitment included 1,017 students; final sample sizes were 204 (Grade 2), 288 (Grade 3), 184 (Grade 4), and 206 (Grade 5). No demographic information was collected. easyCBM® measures included Vocabulary, Common Core State Standards (CCSS) Reading (Grades 3–5 only), Multiple-Choice Reading Comprehension (MCRC), and Passage Reading Fluency (PRF). Comparator measures included the Gates-MacGinitie Reading Tests (Word Knowledge and Reading Comprehension subtests) and DIBELS Oral Reading Fluency (ORF). For logistical reasons, paper-pencil versions of the easyCBM® progress monitoring measures were used. Prior to analysis, assumptions of normality and linearity were evaluated. Because most measures did not meet normality assumptions, Spearman's rank-order correlations (rs) were used for vocabulary and comprehension analyses. Pearson's correlations (r) were used for fluency measures, as assumptions were met. Correlation analyses were conducted separately by grade level to examine the strength and direction of relations between easyCBM® and comparator assessments.

Results

Across grades, correlations between easyCBM® vocabulary and comprehension measures and the Gates-MacGinitie assessments ranged from low to moderate. Vocabulary correlations generally fell in the .30 to .70 range depending on grade. For example, Grade 2 easyCBM® Vocabulary correlated .76 with Gates-MacGinitie Word Knowledge, while Grade 3 correlations were more modest. CCSS and MCRC comprehension measures demonstrated correlations with Gates-MacGinitie Reading Comprehension that ranged from the .40s to the .70s across Grades 3–5, indicating moderate criterion-related validity. In contrast, the strongest validity evidence was observed for fluency measures. Across Grades 2–5, easyCBM® Passage Reading Fluency demonstrated consistently high correlations with DIBELS ORF, with coefficients ranging from the .80s to mid-.90s. These findings indicate strong convergence between the two fluency-based measures. The report notes variability in missing data for the Gates-MacGinitie measures, particularly in Grades 3–5, due to technological and logistical constraints. Despite this limitation, overall findings suggest moderate criterion-related validity evidence for vocabulary and comprehension measures and strong validity evidence for fluency measures. The results support the use of easyCBM® reading assessments for screening and progress monitoring within Grades 2–5, particularly for oral reading fluency.

Table 28. Illustrative Table from Technical Report 1310

Table 3

Correlation Results – Grade 2

Measures		easyCBM® Vocabulary	easyCBM® MCRC	Gates-MacGinitie WK	Gates-MacGinitie RC	DIBELS ORF
easyCBM® Vocabulary	r_s	1	.56**	.76**	.58**	-
	n	233	233	194	176	-
easyCBM® MCRC	r_s	-	1	.61**	.66**	-
	n	-	233	194	176	-
Gates-MacGinitie WK	r_s	-	-	1	.68**	-
	n	-	-	199	181	-
Gates-MacGinitie RC	r_s	-	-	-	1	-
	n	-	-	-	181	-
easyCBM® PRF	r	-	-	-	-	.95**
	n	-	-	-	-	229

Note. r_s = Spearman's rho rank correlation coefficient. r = Pearson's correlation coefficient. **. Correlation is significant at the 0.01 level (2-tailed). easyCBM®-comparator measure coefficients in bold-red fonts. MCRC = Multiple Choice Reading Comprehension, WK = Word Knowledge, RC = Reading Comprehension, PRF = Passage Reading Fluency.

Reference

Lai, C.-F., Alonzo, J., & Tindal, G. (2013). *easyCBM® reading criterion related validity evidence: Grades 2–5 (Technical Report No. 1310)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1103: easyCBM® Reading Criterion-Related Validity Evidence: Oregon State Test 2009–2010 (Anderson, Alonzo, et al., 2011b).

Methods

This report evaluates the criterion-related validity of easyCBM® reading assessments against Oregon's OAKS assessment for grades 3–7. Data were drawn from three Oregon districts. Measures included PRF, MCRC, Vocabulary, and grade 3 WRF. Predictive validity analyses used fall and winter easyCBM® administrations. Concurrent validity analyses used spring administrations. Multiple regression models were conducted, including full-year models and seasonal models. Correlation analyses were conducted by district and full sample. Scatterplots examined classification accuracy relative to OAKS cut scores.

Results

Full annual models explained between 67% and 73% of variance in OAKS scores, exceeding variance explained in the Washington study. At grade 3, the full model accounted for 71% of variance. Removing WRF reduced explained variance by 5%. Vocabulary and MCRC measures frequently served as the strongest unique predictors. Fall and winter models demonstrated strong predictive validity (up to 64% variance explained), while spring models showed strong concurrent validity (up to 66%). Scatterplots indicated that students below the 20th percentile on easyCBM® rarely achieved OAKS proficiency. Classification was more accurate for identifying non-proficient students than proficient students. Overall, findings demonstrate strong criterion-related validity and robust predictive relationships between easyCBM® reading measures and OAKS outcomes.

Table 29. Illustrative Table from Technical Report 1103

District 2 Correlations – Grade 3

		OAKS Rdg	Fall WRF	Fall PRF	Fall MCRC	Fall Voc	Wint WRF	Wint PRF	Wint MCRC	Spr WRF	Spr PRF	Spr MCRC	Spr Voc
OAKS Rdg	Pearson Corr.	1	.651**	.671**	.600**	.708**	.618**	.694**	.601**	.601**	.687**	.682**	.728**
	N	825	775	776	765	762	797	797	789	819	818	814	813
Fall WRF	Pearson Corr.	.651**	1	.919**	.644**	.740**	.916**	.887**	.517**	.875**	.873**	.511**	.688**
	N	775	802	801	784	786	798	798	783	796	795	784	787
Fall PRF	Pearson Corr.	.671**	.919**	1	.652**	.718**	.859**	.906**	.517**	.820**	.898**	.533**	.657**
	N	776	801	802	784	786	798	798	783	796	795	784	787
Fall MCRC	Pearson Corr.	.600**	.644**	.652**	1	.630**	.573**	.647**	.526**	.548**	.626**	.553**	.522**
	N	765	784	784	787	778	782	782	777	783	783	779	781
Fall Voc	Pearson Corr.	.708**	.740**	.718**	.630**	1	.703**	.707**	.559**	.668**	.705**	.581**	.712**
	N	762	786	786	778	787	783	783	774	783	783	774	776
Wint WRF	Pearson Corr.	.618**	.916**	.859**	.573**	.703**	1	.898**	.470**	.905**	.863**	.482**	.679**
	N	797	798	798	782	783	826	826	808	819	818	807	810
Wint PRF	Pearson Corr.	.694**	.887**	.906**	.647**	.707**	.898**	1	.522**	.850**	.913**	.549**	.677**
	N	797	798	798	782	783	826	827	808	820	819	807	810
Wint MCRC	Pearson Corr.	.601**	.517**	.517**	.526**	.559**	.470**	.522**	1	.458**	.515**	.555**	.494**
	N	789	783	783	777	774	808	808	809	808	808	802	802
Spr WRF	Pearson Corr.	.601**	.875**	.820**	.548**	.668**	.905**	.850**	.458**	1	.888**	.479**	.677**
	N	819	796	796	783	783	819	820	808	852	851	837	839
Spr PRF	Pearson Corr.	.687**	.873**	.898**	.626**	.705**	.863**	.913**	.515**	.888**	1	.567**	.670**
	N	818	795	795	783	783	818	819	808	851	851	837	839
Spr MCRC	Pearson Corr.	.682**	.511**	.533**	.553**	.581**	.482**	.549**	.555**	.479**	.567**	1	.599**
	N	814	784	784	779	774	807	807	802	837	837	839	837
Spr Voc	Pearson Corr.	.728**	.688**	.657**	.522**	.712**	.679**	.677**	.494**	.677**	.670**	.599**	1
	N	813	787	787	781	776	810	810	802	839	839	837	841

a. Cannot be computed because at least one of the variables is constant. **Significant at the 0.01 level. *Significant at the 0.05 level.

Reference

Anderson, D., Alonzo, J., & Tindal, G. (2011). *easyCBM reading criterion-related validity evidence: Oregon state test 2009–2010 (Technical Report No. 1103)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1101: easyCBM® Reading Criterion Related Validity Evidence: Washington State Test 2009-2010 (Anderson, Alonzo, et al., 2011a).

This study examined criterion-related validity of easyCBM® reading measures using a large statewide sample of students in Washington across grades 3 through 8. Participants included thousands of students per grade who completed easyCBM assessments (Vocabulary, Multiple Choice Reading Comprehension, and Passage Reading Fluency) administered during fall, winter, and spring benchmark periods. These measures were aligned with reading constructs of comprehension, fluency, and vocabulary knowledge. Data collection procedures involved routine benchmark administrations within school settings, with scores linked to performance on the Washington Measurement of Student Progress (MSP) reading assessment.

Statistical analyses included Pearson correlations to estimate concurrent and predictive validity between easyCBM measures and MSP reading scores. Multiple regression models were also conducted to evaluate the combined predictive power of multiple easyCBM® measures within and across seasons. Model fit indices (R and R^2) and regression coefficients (both unstandardized and standardized) were reported by grade and season, allowing detailed examination of unique and shared contributions of each measure to predicting state test outcomes.

The findings provide strong evidence for the criterion-related validity of easyCBM® reading measures in predicting performance on the Washington MSP reading assessment. Across grades 3 through 8, individual easyCBM® measures demonstrated moderate to strong correlations with MSP scores, typically ranging from approximately .57 to .68 depending on grade, season, and measure. Passage Reading Fluency (PRF) and Multiple Choice Reading Comprehension (MCRC) consistently showed the strongest relationships with MSP outcomes, reflecting their close alignment with reading comprehension constructs.

Regression analyses further indicated that combining measures improved predictive accuracy. Full seasonal models, which included multiple easyCBM® measures within a given benchmark period, produced R^2 values generally in the range of approximately .40 to .55, indicating that a substantial proportion of variance in MSP reading scores could be explained by easyCBM® performance. Models that incorporated data across seasons showed even stronger predictive utility, demonstrating the cumulative value of repeated measurement over time.

Standardized regression coefficients revealed that different measures contributed uniquely to prediction. PRF often had strong contributions in earlier grades, while comprehension measures (MCRC and Vocabulary) became increasingly important in upper grades. This developmental pattern is consistent with theoretical expectations regarding reading development, where fluency supports comprehension in early stages and comprehension processes become more dominant later.

The consistency of findings across grades and seasons supports the stability of the relationships between easyCBM® and MSP outcomes. Additionally, the relatively narrow confidence intervals around correlation coefficients indicate precision in estimates due to large sample sizes. Overall, the results demonstrate that easyCBM® measures provide valid and educationally meaningful indicators of reading proficiency as assessed by statewide accountability tests, supporting their use in screening, progress monitoring, and instructional decision-making contexts.

Table 30. Summary of Key Findings from Technical Report 1101

Grade Range	Measure	Validity Coefficients	Key Findings
Grades 3–8	PRF	.60–.68	Strong predictor of MSP reading, especially in lower grades
Grades 3–8	MCRC	.62–.68	Consistently strong relationship with comprehension outcomes
Grades 3–8	Vocabulary	.57–.65	Moderate to strong predictor; increases in importance in upper grades
Grades 3–8	Full Models	$R^2 \approx .40–.55$	Combined measures substantially improve prediction accuracy

Reference

Alonzo, J., Park, B. J., Tindal, G., & colleagues. (2011). *Technical Report 1101: easyCBM® Reading Criterion-Related Validity Evidence—Washington State Test 2009–2010*. Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1408: Technical Manual: easyCBM® (Anderson et al., 2014).

Methods

Technical evidence for easyCBM® reading measures was gathered from diverse samples spanning Grades K–8. Early literacy samples (Grades K–1) included approximately 2,000 students in the Pacific Northwest, with additional national samples from 76 schools across 26 states. Criterion validity studies for upper-grade reading measures included several hundred to several thousand students per grade, depending on the study. The 2013–2014 norming sample was drawn using a stratified random sampling procedure targeting 500 students per demographic cell (region, race-ethnicity, gender) to reflect national enrollment proportions according to the National Center for Education Statistics Common Core of Data.

The easyCBM® reading assessment system spans multiple measure types: early literacy measures (Letter Names, Letter Sounds, and Phoneme Segmenting, Grades K–1); fluency measures (Word Reading Fluency and Passage Reading Fluency, Grades K–8); and comprehension measures (Vocabulary, Multiple Choice Reading Comprehension [MCRC], and CCSS Reading, Grades 2–8). All measures include three seasonal benchmarks and up to 20–30 progress monitoring forms per grade. CCSS Reading measures assess literal comprehension across three genres: Read to Perform a Task, Informational Text, and Short Literary Text.

Measures were developed through iterative item-writing and review processes involving trained teacher panels and University of Oregon researchers. For fluency measures, piloting occurred across Pacific Northwest schools; CCSS Reading measures were piloted using online national samples. Alignment studies employed teacher raters who independently judged item-standard correspondence, with inter-rater reliability evaluated across multiple sessions. Criterion validity data were collected by administering easyCBM® benchmarks and external criterion tests (e.g., DIBELS, Stanford Achievement Test-10 [SAT-10], state tests, Oregon Assessment of Knowledge and Skills [OAKS]) within the same school year.

Rasch modeling was applied to all discrete-item measures to calibrate item difficulty and evaluate fit (Mean Square Outfit; acceptable range 0.50–1.50). Because Passage Reading Fluency measures lack discrete items, Analysis of Variance (ANOVA) was used to evaluate test form equivalence, supplemented by Flesch-Kincaid readability indices. Reliability was assessed via Pearson's r (alternate form and test-retest), Cronbach's alpha (internal consistency), and generalizability theory (G-coefficients). Criterion validity analyses used Pearson's r, Spearman's rho (for non-normal distributions), simple and multiple linear regression, and diagnostic efficiency statistics (sensitivity, specificity, AUC). Construct validity was examined through confirmatory factor analysis (CFA) using one-, two-, and three-factor models.

Results

Rasch-based item development was employed across all three early literacy measures to ensure test form equivalence. Alternate form reliability was consistently strong: for Letter Names, same-day alternate form correlations ranged from .87 to .90 for Grade K and .85 to .90 for Grade 1; test-retest correlations ranged from .79 to .82. Letter Sounds alternate form reliability ranged from .76 to .92, and generalizability theory G-coefficients ranged from .87 to .95, indicating high measurement consistency. Phoneme Segmenting showed similarly robust alternate form reliability, with G-coefficients exceeding .80 across forms.

Criterion validity for early literacy measures was demonstrated through multiple studies. Letter Names showed Spearman's rho correlations of .86 (Grade K) and .80 (Grade 1) with DIBELS Letter Naming Fluency. Combined early literacy measures (Letter Names, Letter Sounds, Phoneme Segmenting) predicted 35–58% of variance in Kindergarten SAT-10 scores and 49–56% of variance in Grade 1 SAT-10 Word Reading Fluency performance. Letter Sounds showed moderate concurrent correlations with DIBELS measures (.55 with ISF at Grade K; .58 with NWF at Grade 1). Confirmatory factor analyses supported a single latent "Reading" factor at both grade levels, with Letter Names factor loadings ranging from the .80s to .90s (Kindergarten) and .90s (Grade 1), confirming strong construct validity.

Word Reading Fluency (WRF) and Passage Reading Fluency (PRF) measures were developed with ANOVA-based equating to ensure form equivalence, supplemented by Rasch modeling for WRF. Both measures demonstrated strong alternate form and test-retest reliability, generally exceeding .80 across grades. PRF generalizability studies indicated that the passage facet (i.e., differences across passages) was the largest source of measurement variance after the student facet, which was addressed through the ANOVA-based equating process.

Criterion validity for fluency measures was strong and consistent. Predictive and concurrent correlations with external criteria (state tests, DIBELS Oral Reading Fluency, SAT-10) were generally in the .60–.85 range across grades. Fluency measures consistently accounted for substantial proportions of variance in year-end reading outcomes. Construct validity analyses placed WRF and PRF within a unidimensional reading factor at Grades K–2, and as part of a fluency factor distinct from comprehension at Grade 5, reflecting expected developmental differentiation. CFA factor loadings were consistently in the moderate-to-strong range (.50–.85).

The Multiple-Choice Reading Comprehension (MCRC) and CCSS Reading measures were developed using Rasch modeling to ensure form equivalence across all 20 progress monitoring and 3 benchmark forms per grade. Internal consistency (Cronbach's alpha) for CCSS Reading was investigated by Guerreiro et al. (2014) with strong results reported across Grades 3–8. Criterion validity studies showed that easyCBM® reading measures at Grades 2–5 had significant predictive and concurrent correlations with criterion measures (Lai, Alonzo, & Tindal, 2013), and the CCSS Reading measures demonstrated strong alignment with grade-level literacy standards.

Construct validity analyses for CCSS Reading (Grades K–5) used CFA to evaluate competing factor structures. For Grades K–2, the three-factor model (phonological awareness, fluency, comprehension) fit the data best, suggesting that these three components of reading are distinguishable at early grades. For Grades 3 and 4, the one-factor model without testlet effects provided the best fit, indicating a general reading factor. For Grade 5, the two-factor model without a testlet effect fit best, with factors representing (a) fluency (WRF and PRF) and (b) comprehension (CCSS and MCRC). For Grades 3–5 CCSS-only models, the one-factor model also showed the best fit, with item loadings on the latent reading comprehension factor ranging from .60–.80 (Grade 3), .50–.70 (Grade 4), and .10–.70 with a median of .58 (Grade 5).

Across all reading measure types and grade levels, the easyCBM® reading assessments demonstrated strong technical adequacy. Reliability evidence was consistently acceptable to strong, with alternate form correlations, test-retest coefficients, and Cronbach's alpha values typically meeting or exceeding the .80 threshold. Criterion validity evidence was robust, with moderate-to-strong correlations across multiple external benchmarks and state assessments. Construct validity evidence supported a unidimensional reading factor at most grade levels, with developmental differentiation into subfactors (fluency vs. comprehension) emerging in upper elementary grades, consistent with theoretical models of reading development. The alignment of CCSS Reading items to literacy standards was generally strong, providing content validity support for their intended use as RTI screening and progress monitoring tools.

Table 31. Summary of Results from Technical Report 1408

Measure	Analysis Type	Metric	Result
Letter Names (K–1)	Alternate Form Reliability	Pearson's r (same-day)	.87–.90 (K); .85–.90 (Gr. 1)
Letter Names (K–1)	Criterion Validity	Spearman's ρ vs. DIBELS LNF	.86 (K); .80 (Gr. 1)
Letter Sounds (K–1)	Generalizability	G-coefficient	.87–.95
Letter Sounds (K–1)	Concurrent Validity	ρ vs. DIBELS ISF/NWF	.55 (K); .58 (Gr. 1)
Early Literacy Battery (K–1)	Predictive Validity	R^2 vs. SAT-10	.35–.58 (K); .49–.56 (Gr. 1)
Early Literacy (K–1)	Construct Validity (CFA)	Factor Loadings on Reading	.80s–.90s (K); .90s (Gr. 1)
Fluency Measures (K–8)	Reliability	Alternate Form / Test-retest r	Generally ≥.80 across grades
Fluency Measures (K–8)	Criterion Validity	r vs. State Tests / DIBELS ORF	.60–.85 across grades
CCSS Reading (K–2)	Construct Validity (CFA)	Best-fit Model	Three-factor (phonological, fluency, comprehension)
CCSS Reading (Gr. 3–4)	Construct Validity (CFA)	Best-fit Model	One-factor (general reading)
CCSS Reading (Gr. 5)	Construct Validity (CFA)	Best-fit Model	Two-factor (fluency; comprehension)
CCSS Reading (Gr. 3–5)	Construct Validity (CFA)	Item Loadings (CCSS-only model)	.60–.80 (Gr. 3); .50–.70 (Gr. 4); Mdn .58 (Gr. 5)
CCSS Reading (Gr. 3–8)	Internal Consistency	Cronbach's α	Strong across grades

Reference

Anderson, D., Alonzo, J., Tindal, G., Farley, D., Irvin, P. S., Lai, C. F., Saven, J. L., & Wray, K. A. (2014). *Technical manual: easyCBM (Technical Report # 1408).* Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Summary of Technical Report 1702: An Update to Compiled ORF Norms (Hasbrouck & Tindal, 2017).

Methods

This technical report presents an updated compilation of oral reading fluency (ORF) norms for students in Grades 1 through 6. The purpose of the study was to revise previously published compiled norms from 1992 and 2006 to reflect more current reading performance data and assessment practices. Unlike early ORF norm development, which relied on locally created passages and district-level benchmarks, this update draws upon large-scale datasets from widely used, commercially available curriculum-based measurement (CBM) systems.

Data were obtained from three ORF assessments: DIBELS 6th Edition, DIBELS Next, and easyCBM®. Rather than collecting raw student-level data directly from schools, the authors extracted percentile rank information from technical reports produced by the publishers of each assessment. These reports included norms derived from very large national samples. The total number of ORF scores represented across the three assessments exceeded 6.8 million. DIBELS datasets represented complete population samples for their respective years, while easyCBM® norms were based on stratified random sampling to ensure demographic representation across region, gender, and ethnicity.

A master dataset was constructed that included grade level, testing season (fall, winter, spring), percentile rank, assessment source, and words correct per minute (WCPM). For each grade, season, and percentile rank, the corresponding WCPM values from the three assessments were averaged to produce compiled norms. Consistent with earlier reports, only the 10th, 25th, 50th, 75th, and 90th percentiles were reported. The resulting norms were then compared to the 2006 compiled norms to evaluate change over time.

Results

The updated 2017 compiled ORF norms generally show modest increases in WCPM scores compared to the 2006 norms, particularly in Grades 1 through 5. In these grade levels, most percentile benchmarks increased across fall, winter, and spring administrations. Average gains across percentiles ranged from approximately 6 to 12 WCPM depending on grade, with the largest increases occurring in Grade 3. Lower percentile levels, especially the 10th and 25th percentiles, tended to show the greatest improvements, suggesting stronger performance growth among lower-performing students relative to earlier benchmarks. Grade 6 demonstrated a different pattern. Although many percentile scores increased slightly, some remained unchanged and a few decreased marginally. Overall average gains in Grade 6 were smaller than in earlier grades. Across all six grades and percentile levels combined, the average increase from 2006 to 2017 was approximately five WCPM, a change that falls within expected developmental growth ranges reported in prior research.

The updated norms provide evidence of general stability in ORF performance across nearly three decades of reading instruction, while also reflecting incremental improvements in reading fluency outcomes. The authors conclude that compiled norms remain valuable for screening and progress monitoring because they reduce the risk of setting benchmarks too low in underperforming local contexts. These updated benchmarks allow educators to interpret student performance relative to broad, aggregated national data rather than relying solely on local comparisons.

Table 32. Illustrative Table from Technical Report 1702

%iles	Grade 1	F	W	S	Grade 2	F	W	S
90	2017		97	116	2017	111	131	148
90	2006		81	111	2006	106	125	142
	Difference		16	5	Difference	5	6	6
75	2017		59	91	2017	84	109	124
75	2006		47	82	2006	79	100	117
	Difference		12	9	Difference	5	9	7
50	2017		29	60	2017	50	84	100
50	2006		23	53	2006	51	72	89
	Difference		6	7	Difference	-1	12	11
25	2017		16	34	2017	36	59	72
25	2006		12	28	2006	25	42	61
	Difference		4	6	Difference	11	17	11
10	2017		9	18	2017	23	35	43
10	2006		6	15	2006	11	18	31
	Difference		3	3	Difference	12	17	12

%iles	Grade 3	F	W	S	Grade 4	F	W	S
90	2017	134	161	166	2017	153	168	184
90	2006	128	146	162	2006	145	166	180
	Difference	6	15	4	Difference	8	2	4
75	2017	104	137	139	2017	125	143	160
75	2006	99	120	137	2006	119	139	152
	Difference	5	17	2		6	4	8
50	2017	83	97	112	2017	94	120	133
50	2006	71	92	107	2006	94	112	123
	Difference	12	5	5	Difference	0	8	10
25	2017	59	79	91	2017	75	95	105
25	2006	44	62	78	2006	68	87	98
	Difference	15	17	13		7	8	7
10	2017	40	62	63	2017	60	71	83
10	2006	21	36	48	2006	45	61	72
	Difference	19	26	15	Difference	15	10	11

Reference

Hasbrouck, J., & Tindal, G. (2017). *An update to compiled ORF norms (Technical Report No. 1702)*. Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 1701: Oral Reading Fluency: Outcomes from 30 Years of Research (Tindal, 2017).

Methods

This technical report synthesizes approximately 30 years of research on Oral Reading Fluency (ORF) and its role within curriculum-based measurement (CBM) and large-scale accountability systems. Rather than presenting new experimental data, the report uses a research synthesis approach, drawing on empirical studies, meta-analytic findings, and technical adequacy research conducted across multiple decades. The purpose was to examine the evolution of ORF as a measurement tool and to evaluate the evidence base supporting its reliability, validity, and instructional utility.

The report reviews foundational CBM studies from the 1980s and 1990s that established ORF as a sensitive indicator of reading competence. It then integrates subsequent research examining alternate-form reliability, growth modeling, decision rules, and criterion-related validity with high-stakes state assessments. Attention is given to methodological variations across studies, including differences in passage selection, administration time, scoring conventions (words correct per minute), and equating procedures. The synthesis also addresses criticisms of ORF, including concerns about construct representation, overemphasis on speed, and whether ORF sufficiently captures comprehension. The report evaluates empirical evidence related to these concerns by examining correlations between ORF and comprehension measures, growth trajectories across grade levels, and predictive relationships with reading outcomes.

Results

Across three decades of research, findings consistently demonstrate strong technical adequacy for ORF as a screening and progress monitoring measure. Studies reviewed in the report indicate high alternate-form reliability and strong correlations between ORF and broader reading proficiency measures, including comprehension-based assessments. ORF scores have repeatedly shown substantial predictive relationships with state reading tests, particularly in elementary grades.

The report highlights that ORF functions as a robust general indicator of reading competence, especially in early and middle elementary grades where decoding and fluency are tightly linked to comprehension development. Growth analyses across studies indicate that ORF is sensitive to instructional change and responsive to intervention effects, making it particularly useful in multi-tiered systems of support (MTSS). While acknowledging that ORF does not directly measure deep comprehension processes, the report concludes that it serves as a highly efficient and empirically supported proxy for overall reading proficiency. Limitations become more apparent in upper grades, where comprehension demands diversify and the strength of ORF–comprehension correlations may attenuate. Overall, the synthesis supports continued use of ORF as a technically sound, efficient, and instructionally meaningful measure within screening and progress monitoring systems.

Table 33. Illustrative Table from Technical Report 1701

Table 1

Comparison of Studies on Research Variables Referenced in Studying Growth of Reading Fluency

Authors (date)	Grades	Students	Measures	N-Measures	Slope Calculation	Averages and Growth – WCPM
Marston, Lowry, Deno, & Mirkin (1981)	Grades (n): 1 (13), 2 (9), 3 (10), 4 (7), 5 (7), 6 (9)	Students from a small Midwest city	Third grade basal reading series: Allyn-Bacon, Ginn 720, Houghton Mifflin	Three administrations: Fall, winter, and spring	Plot of raw scores and percentage of change; significance tests of differences	Grades / Fall / Winter / Spring 1 / 18.1 / 31.1 / 45.7 2 / 73.2 / 101.1 / 127.8 3 / 108.3 / 123.6 / 136.2 4 / 125.4 / 131.7 / 155.3 5 / 125.7 / 147.3 / 161.1 6 / 142.9 / 176.7 / 182.8
Tindal, Germann, Marston, & Deno (1983)	Grades (n): 1 (5), 2 (13), 3 (17), 4 (22), 5 (18), 6 (21)	Students from six districts referred, assessed, and eligible for special education	Curricula in use in the school district	Three administrations: Fall, winter, and spring	Change in raw score and in discrepancy from general education	Grades / Fall / Winter / Spring 1 / 6.8 / 14.3 / 9.6 2 / 5.5 / 16.2 / 23.6 3 / 20.5 / 36.6 / 41.3 4 / 31.0 / 50.3 / 52.9 5 / 59.1 / 72.9 / 79.1 6 / 59.5 / 65.9 / 66.8
Tindal, Germann, & Deno (1983)	Grades (n): 1 (276) 2 (284) 3 (302) 4 (294) 5 (315) 6 (328)	Students randomly sampled from six districts in Pine County	Two passages sampled from basal reading curriculum	Three administrations: Fall, winter, and spring	Change in raw scores	Grades / Fall / Winter / Spring 1 / 5 / 63 / 75 2 / 35 / 83 / 93 3 / 67 / 89 / 108 4 / 98 / 111 / 128 5 / 121 / 120 / 138 6 / 123 / 126 / 134
Fuchs, Deno, & Mirkin (1984)	Grades 3-5: 64 students in (DBPM) 77 students in no DBPM	All students were 'handicapped'	3rd grade passage reading test from Ginn 720	Pre-post (unknown: sometime between Nov. and May)	Pre-post difference @ 28 weeks	Condition* / Pre / Post Experimental* / 41.6 / 70.2 Contrast / 51.5 / 51.3 *DBPM versus none*

Reference

Tindal, G. (2017). *Oral reading fluency: Outcomes from 30 years of research (Technical Report No. 1701).* Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Summary of Technical Report 0030: District Reading Assessments, Spring 2004 Administration (Alonzo & Tindal, 2004e).

Methods

This report documents district-wide administration of a comprehensive K–8 reading assessment system in spring 2004, with pilot data from grade 9. The assessment system was developed collaboratively between district educators and University of Oregon researchers to meet eight design criteria including psychometric soundness, use of authentic literature, progress monitoring capability, and alignment with state tests. Subtests varied by grade level and included Letter Naming, Letter Sounds, Phoneme Segmentation, High Frequency Words, Word Reading, Oral Reading Fluency (ORF), Multiple Choice (MC) Comprehension, Vocabulary, and Maze Comprehension. Administration was conducted by trained retired or substitute teachers. Data were captured using optical character recognition scan sheets and analyzed at the university. Analyses included descriptive statistics (means, SDs, percentile ranks), district norm development, and Rasch-based item response theory (IRT) modeling for multiple choice and other selected response measures. Misfit statistics, item redundancy, ceiling effects, and discrimination patterns were evaluated. Recommendations were generated for item revision and form abbreviation.

Results

District norm tables were produced for each grade level. Early literacy measures showed expected developmental progression, with strong discrimination across percentile bands. ORF measures demonstrated acceptable discrimination and reliable differentiation across ability levels. Vocabulary measures at grades 3 through 6 exhibited ceiling effects, limiting sensitivity among higher-performing students. As a result, the committee recommended increasing item difficulty. MC Comprehension measures generally demonstrated acceptable Rasch model fit. However, redundant items were identified in multiple grade levels, and several poorly fitting items were recommended for removal or revision. Grade 9 measures, piloted for the first time, showed appropriate discrimination but required shortening to align with district testing time constraints. Specific items were flagged for redundancy or misfit. The district accepted recommendations to revise poorly fitting items, increase vocabulary difficulty at several grade levels, rewrite Grade 8 ORF passages to address readability imbalances, and eliminate open-ended comprehension items due to limited incremental information. Overall, the system was deemed technically sound and highly informative for screening, intervention planning, and longitudinal tracking.

Reference

Alonzo, J., & Tindal, G. (2004). *District reading assessments, spring 2004 administration (Technical Report No. 0030).* Eugene, OR: University of Oregon, Behavioral Research and Teaching.

Appendix A: Technical Report Table and Figure Titles

Appendix B: Guide to Spreadsheet Technical Report Value Displays

See Riverside Insights or BRT to access exact values for TR Summaries
2603-VK8R_ValidityReadingTables.xlsx

- 1003
- 1004
- 1005
- 1309
- 0922
- 1801
- 1403
- 1805
- 1806
- 1807
- 2401
- 0910
- 1106
- 1107
- 1108
- 1109
- 1302
- 1304
- 1305
- 0025
- 0022
- 0027
- 0028
- 0024
- 0023
- 1310
- 1103
- 1101
- 1408…Not summarized in a spreadsheet
- 1702…Not summarized in a spreadsheet
- 1701…Not summarized in a spreadsheet
- 0030…Not summarized in a spreadsheet

Technical Report References

Alonzo, J., & Anderson, D. (2018a). *Classification accuracy of the easyCBM Kindergarten – Grade 2 reading measures (Technical Report # 1801)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., & Anderson, D. (2018b). *Supplementary report on easyCBM MCRC Measures: A Follow-up to previous technical report (Technical Report # 1807)*. Eugene OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., & Anderson, D. (2018c). *Supplementary report on easyCBM PRF measures: A Follow-up to previous technical reports (Technical Report #1806)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Park, B. J., & Tindal, G. (2013a). *An examination of the internal structures of the easyCBM CCSS reading measures (Technical Report # 1304)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Park, B. J., & Tindal, G. (2013b). *An examination of the internal structures of the Grades K-5 easyCBM CCSS reading measures: A Construct validity study (Technical Report # 1305)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., Park, B. J., & Tindal, G. (2013c). *Examining the internal structure of the easyCBM reading measures, Grades K-5 (Technical Report # 1302)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., & Tindal, G. (2004a). *Analysis of reading fluency and comprehension measures for first-grade students (Technical Report # 25)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., & Tindal, G. (2004b). *Analysis of reading fluency and comprehension measures for fourth-grade students (Technical Report # 27)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., & Tindal, G. (2004c). *Analysis of reading fluency and comprehension measures for seventh-grade students (Technical Report # 23)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., & Tindal, G. (2004d). *Analysis of reading fluency and comprehension measures for sixth-grade students (Technical Report # 24)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Alonzo, J., & Tindal, G. (2004e). *District reading assessments, spring 2004 administration (Technical Report # 30)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

American Educational Research Association, American Psychological Association, & Education, N. C. o. M. i. (2014). *Publication manual* (7th ed.). Author.

Anderson, D., Alonzo, J., & Tindal, G. (2011a). *easyCBM reading criterion related to validity evidence: Washington state test 2009-2010 (Technical Report # 1101)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Alonzo, J., & Tindal, G. (2011b). *easyCBM reading criterion related validity evidence: Oregon state test 2009-2010 (technical Report # 1103)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Alonzo, J., Tindal, G., Farley, D., Irvin, P. S., Lai, C. F., Saven, J. L., & Wray, K. A. (2014). *Technical manual: easyCBM (Technical Report # 1408)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Anderson, D., Park, B. J., Irvin, P. S., Alonzo, J., & Tindal, G. (2011). *Diagnostic efficiency of easyCBM reading: Washington state (Technical Report # 1107)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Beghetto, R., & Tindal, G. (2004). *Analysis of reading fluency and comprehension measures for fifth grade students (Technical Report # 28)*. Behavioral Research and Teaching, University of Oregon.

Guerreiro, M., Alonzo, J., & Tindal, G. (2014). *Internal consistency of the easyCBM CCSS reading measures grades K-8 (Technical Report # 1407)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Hasbrouck, J., & Tindal, G. (2017). *An update to compiled ORF norms (Technical Report # 1702)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Irvin, P. S., Park, B. J., Anderson, D., Alonzo, J., & Tindal, G. (2011). *A cross-validation of easyCBM reading cut scores in Washington: 2009-2010 (Technical Report # 1109)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Jamgochian, E. M., Park, B. J., Nese, J. F. T., Lai, C. F., Sáez, L., Anderson, D., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM Grade 2 reading measures (Technical Report # 1004)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Ketterlin-Geller, L. R., & Tindal, G. (2004). *Analysis of reading fluency and comprehension measures for 3rd-grade students (Technical Report # 22)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Lai, C. F., Alonzo, J., & Tindal, G. (2013a). *easyCBM reading criterion related validity evidence: Grades 2-5 (Technical Report # 1310)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Lai, C. F., Alonzo, J., & Tindal, G. (2013b). *easyCBM reading criterion related validity evidence: Grades K-1 (Technical Report # 1309)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Lai, C. F., Nese, J. F. T., Jamgochian, E. M., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM primary-level reading measures (grades k-1), 2009-2010 version (technical report 1003)*.

Park, B. J., Anderson, D., Irvin, P. S., Alonzo, J., & Tindal, G. (2011). *Diagnostic efficiency of easyCBM reading: Oregon (Technical Report # 1106)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Park, B. J., Irvin, P. S., Anderson, D., Alonzo, J., & Tindal, G. (2011). *A cross-validation of easyCBM reading cut scores in Oregon: 2009-2010 (Technical Report # 1108)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Sáez, L., Park, B. J., Nese, J. F. T., Jamgochian, E. M., Lai, C. F., Anderson, D., Kamata, A., Alonzo, J., & Tindal, G. (2010). *Technical adequacy of the easyCBM reading measures (Grades 3-7), 2009-2010 version (Technical Report # 1005)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Tindal, G. (2017). *Oral reading fluency: Outcomes from 30 years of research (Technical Report # 1701)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Tindal, G., Nese, J. F., & Alonzo, J. (2009a). *Criterion-related evidence using easyCBM® reading measures and student demographics to predict state test performance in Grades 3-8 (Technical Report # 0910)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Tindal, G., Nese, J. F., & Alonzo, J. (2009b). *Hierarchical linear modeling of passage reading fluency growth as a function of student characteristics (Technical Report # 0922)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Tindal, G., & Nese, J. F. T. (2024). *Criterion validity and classification accuracy of easyCBM: Grades 3-8. (Technical Report No. 2401).* Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Wray, K., Lai, C. F., Sáez. L., Alonzo, J., & Tindal, G. (2014). *easyCBM beginning reading measures: Grades K-1 alternate form reliability and criterion validity with the SAT-10 (Technical Report # 1403)*. Eugene, OR: Behavioral Research and Teaching, University of Oregon.